

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE ENGENHARIAS, CIÊNCIAS AGRÁRIAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

MATEUS FELIPE RIBEIRO

DA FLORESTA SECUNDÁRIA À AGRICULTURA: MACHINE LEARNING E
BIOINFORMÁTICA NA AVALIAÇÃO DOS EFEITOS DO USO DO SOLO SOBRE O
MICROBIOMA EDÁFICO NO ESTADO DO PARANÁ

PONTA GROSSA

2026

MATEUS FELIPE RIBEIRO

DA FLORESTA SECUNDÁRIA À AGRICULTURA: MACHINE LEARNING E
BIOINFORMÁTICA NA AVALIAÇÃO DOS EFEITOS DO USO DO SOLO SOBRE O
MICROBIOMA EDÁFICO NO ESTADO DO PARANÁ

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada, curso de Mestrado em Computação Aplicada da Universidade Estadual de Ponta Grossa, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Orientação: Prof. Dr. Rafael Mazer Etto.

PONTA GROSSA

2026

R484 Ribeiro, Mateus Felipe

Da floresta secundária à agricultura: machine learning e bioinformática na avaliação dos efeitos do uso do solo sobre o microbioma edáfico no estado do Paraná / Mateus Felipe Ribeiro. Ponta Grossa, 2026.

100 f.

Dissertação (Mestrado em Computação Aplicada - Área de Concentração: Computação para Tecnologias em Agricultura), Universidade Estadual de Ponta Grossa.

Orientador: Prof. Dr. Rafael Mazer Etto.

1. Microbiota do solo. 2. Uso do solo. 3. Aprendizado de máquina. 4. Bioindicadores. 5. Redes ecológicas. I. Etto, Rafael Mazer. II. Universidade Estadual de Ponta Grossa. Computação para Tecnologias em Agricultura. III.T.

CDD: 006.31



UNIVERSIDADE ESTADUAL DE PONTA GROSSA
Av. General Carlos Cavalcanti, 4748 - Bairro Uvaranas - CEP 84030-900 - Ponta Grossa - PR - <https://uepg.br>

TERMO

TERMO DE APROVAÇÃO

Mateus Felipe Ribeiro

“DA FLORESTA SECUNDÁRIA À AGRICULTURA: MACHINE LEARNING E BIOINFORMÁTICA NA AVALIAÇÃO DOS EFEITOS DO USO DO SOLO SOBRE O MICROBIOMA EDÁFICO NO ESTADO DO PARANÁ”

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:

Prof. Dr. Rafael Mazer Etto (UEPG/PR - Presidente)

Prof. Dr. Adriano Ferrasa (UEPG/PR)

Dr. Ederson da Conceição Jesus (EMBRAPA)

Ponta Grossa, 05 de fevereiro de 2026.



Documento assinado eletronicamente por **Adriano Ferrasa, Professor(a)**, em 23/02/2026, às 12:10, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Rafael Mazer Etto, Professor(a)**, em 23/02/2026, às 13:48, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Ederson da Conceição Jesus, Usuário Externo**, em 23/02/2026, às 18:32, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida no site <https://sei.uepg.br/autenticidade> informando o código verificador **2996968** e o código CRC **348581CB**.

AGRADECIMENTOS

Agradeço ao meu orientador, Dr. Rafael Mazer Etto por toda a acolhida, suporte, ensinamentos, apoio, confiança e incentivo na realização deste trabalho.

Agradeço aos professores da UEPG que me permitiram uma formação rica interdisciplinar, permitindo assim, que eu realizasse um trabalho que percorresse diversas áreas do conhecimento, como deve ser na ciência.

Agradeço ao Programa de Pós-Graduação em Computação Aplicada, por meio dos laboratórios InfoAgro, ao Programa de Pós-Graduação em Agronomia, por meio do Laboratório de Biologia Molecular (LABMOM) pela disponibilização de suas estruturas para a integração com demais acadêmicos de diferentes áreas de estudo, contribuindo na interdisciplinaridade.

Agradeço à minha família, em especial aos meus pais João Carlito Ribeiro e Luciane Rodrigues Ribeiro, a minha irmã Ester Sara Ribeiro.

Agradeço ao meu amigo Douglas Tomachewski por ser referência na carreira acadêmica.

Agradeço aos professores que gentilmente aceitaram o convite para compor a banca de qualificação e defesa.

Agradeço aos meus colegas do Programa de Pós-Graduação em Computação Aplicada pelo apoio, conversas e ensinamentos na área computacional.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

O solo abriga uma comunidade microbiana complexa e essencial para a manutenção dos serviços ecossistêmicos e da sustentabilidade agrícola. A conversão de ecossistemas naturais em áreas agrícolas altera drasticamente as propriedades edáficas e a microbiota do solo. Para compreender essas mudanças, é essencial integrar a ecologia tradicional a abordagens analíticas de alta resolução. Este estudo objetivou propor um fluxo analítico, unindo estatística multivariada e aprendizado de máquina, para modelar a estruturação bacteriana e inferir seus direcionadores edáficos. Sob a hipótese de que modelos preditivos revelam assinaturas ecológicas complexas do manejo antrópico, utilizaram-se dados edáficos integrados ao sequenciamento de rRNA 16S de quatro localidades do Paraná. Empregaram-se ferramentas de bioinformática (DADA2), Partição de Variação e algoritmos computacionais para avaliar a influência do uso do solo na microbiota. Os modelos capturaram assinaturas ecológicas específicas de cada manejo, demonstrando que os atributos biológicos são os principais direcionadores da composição microbiana. Evidenciou-se forte acoplamento entre sistemas agrícolas e filos copiotróficos (*Pseudomonadota*), contrastando com a retenção de oligotróficos sensíveis (*Acidobacteriota* e *Verrucomicrobiota*) na Floresta Secundária (FS). A análise topológica de redes indicou que a FS possui maior complexidade e forte acoplamento bioquímico, sustentando a maior diversidade de ASVs detectáveis. Considerando o limite de detecção nos sistemas agrícolas, a floresta atua como reservatório e ecótono dinâmico. Contudo, sua eficácia funcional na paisagem depende do dimensionamento espacial, visto que fragmentos diminutos sofrem efeito de borda e homogeneização biológica. Conclui-se que integrar aprendizado de máquina e estatística multivariada é uma abordagem robusta para identificar padrões de estabilidade ecológica. Infere-se que a resiliência dos agroecossistemas avaliados atrela-se à manutenção de práticas conservacionistas e remanescentes florestais adequados. O trabalho fornece um fluxo analítico essencial para o futuro desenvolvimento de bioindicadores edáficos.

Palavras-chave: Microbiota do solo; Uso do solo; Aprendizado de Máquina; Bioindicadores; Agroecossistemas; Redes ecológicas.

ABSTRACT

The soil is home to a complex microbial community that is essential for maintaining ecosystem services and agricultural sustainability. The conversion of natural ecosystems into agricultural areas drastically alters edaphic properties and soil microbiota. To understand these changes, it is essential to integrate traditional ecology with high-resolution analytical approaches. This study aimed to propose an analytical workflow, combining multivariate statistics and machine learning, to model bacterial structuring and infer its edaphic drivers. Under the hypothesis that predictive models reveal complex ecological signatures of anthropogenic management, edaphic data integrated with 16S rRNA sequencing from four locations in the state of Paraná were used. Bioinformatics tools (DADA2), Variation Partitioning, and computational algorithms were employed to evaluate the influence of land use on the microbiota. The models captured specific ecological signatures for each management type, demonstrating that biological attributes are the main drivers of microbial composition. A strong coupling was evidenced between agricultural systems and copiotrophic phyla (*Pseudomonadota*), contrasting with the retention of sensitive oligotrophs (*Acidobacteriota* and *Verrucomicrobiota*) in the Secondary Forest (SF). Topological network analysis indicated that the SF exhibits greater complexity and strong biochemical coupling, sustaining the highest diversity of detectable ASVs. Considering the computational detection limit in agricultural systems, the forest acts as a reservoir and a dynamic ecotone. However, its functional effectiveness in the landscape depends on spatial scaling, since diminutive fragments suffer from edge effects and biological homogenization. It is concluded that integrating machine learning and multivariate statistics is a robust approach for identifying patterns of ecological stability. It is inferred that the resilience of the evaluated agroecosystems is closely linked to the maintenance of conservation practices and adequate forest remnants. This work provides an essential analytical workflow for the future development of edaphic bioindicators.

Keywords: Soil microbiota; Land use; Machine learning; Bioindicators; Agroecosystems; Ecological networks.

LISTA DE FIGURAS

Figura 1 - Locais de estudo no estado do Paraná	22
Figura 2 - Fluxograma Método Processamento 16S rRNA.....	27
Figura 3 - Análise de qualidade biblioteca de sequências RNA.....	28
Figura 4 - Curvas SRS para cada Tratamento	31
Figura 5 - Corte amostras através do SRS	31
Figura 6 - Frequência de ASVs nas Amostras por SRS	32
Figura 7 - Análise gráfica normalização da abundância de ASVs	33
Figura 8 - Análise gráfica filtrando abundância de ASVs para apenas domínio Bacteria	33
Figura 9 - Classificação taxonômica das comunidades bacterianas do solo sob diferentes manejos.....	42
Figura 10 - Diversidade alfa das comunidades bacterianas.....	43
Figura 11 - Análise de Coordenadas Principais (PCoA)	46
Figura 12 - Ordenação Não-Métrica NMDS	46
Figura 13 - Ordenação Restringida dbRDA	47
Figura 14 - Análise de Componentes Principais (PCA).....	49
Figura 15 - Análise de Partição de Variação através de Diagramas de Venn	50
Figura 16 - Gráficos de importância de variáveis derivados de modelos de Random Forest ..	52
Figura 17 - Gráficos de importância de variáveis derivados de modelos de XGBoost.....	52
Figura 18 - Gráficos de importância de variáveis derivados de modelos de GBM.....	53
Figura 19 - Rede de correlação de Spearman para o tratamento Plantio Direto	58
Figura 20 - Rede de correlação de Spearman para o tratamento Floresta Secundária	59
Figura 21 - Rede de correlação de Spearman para o tratamento de Pastagem.....	61
Figura 22 - Análise de Espécies Indicadoras.....	66
Figura 23 - Filos Espécies Indicadoras em cada Tratamento	68
Figura 24 - Filos Espécies Não-Indicadoras em cada Tratamento	69
Figura 25 - Relação Amplitude Nicho Abundância Levins' B Index	70
Figura 26 - Distribuição Amplitude Nicho Levins' B Index.....	70
Figura 27 - Mapa de calor da abundância relativa de vias metabólicas entre tratamentos.....	73

LISTA DE QUADROS

Quadro 1 - Trabalhos relacionados a análises estatísticas a partir dos dados de 16s de microrganismos do solo.....	19
Quadro 2 - Descrição dos locais de estudo.....	22
Quadro 3 - Composição do conjunto de dados segundo a localidade de amostragem, município e as classes de uso e manejo do solo	23
Quadro 4 – Representação dos nomes de pontos de coleta	30

LISTA DE TABELAS

Tabela 1 - Comparação valores preditos por VarPart, RF, XGBoost e GB para atributos biológicos, químicos e físicos e residual	55
Tabela 2 - Valores explicativos e de validação dos algoritmos de aprendizado de máquina...	56

LISTA DE ABREVIATURAS E SIGLAS

AEI	Análise de Espécie Indicadora
AM	Aprendizado de Máquina
ANOVA	Análise de Variância
ASV	Variante de Sequência Amplicon
DADA2	Algoritmo Divisivo de Descorreção de Amplicons
IVI	Índice de Valor Indicador
NCBI	Centro Nacional de Informação em Biotecnologia
OTU	Unidade Taxonômica Operacional
PCA	Análise de Componentes Principais
PCoA	Análise de Coordenadas Principais
PCR	Reação em Cadeia da Polimerase
PV	Partição de Variação
QIIME	Percepções Quantitativas Sobre Ecologia Microbiana
RF	Algoritmo Floresta Aleatória
rRNA	Ácido Ribonucleico Ribossomal
SILVA	Banco de Dados Abrangente de Genes de rRNA
SRS	Escalonamento com Subamostragem Ranqueada
XGBoost	Algoritmo Aumento de Gradiente Extremo
GB	Algoritmo Aumento de Gradiente
FS	Floresta Secundária
PD	Plantio Direto
PA	Pastagem
PC	Plantio Convencional
SS	Sistema Silvipastoril

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO DA LITERATURA	14
2.1	EVOLUÇÃO DO USO DO GENE 16S RRNA EM ANÁLISES MICROBIANAS	14
2.2	SAÚDE DO SOLO E INDICADORES DE QUALIDADE	15
2.3	SISTEMAS DE MANEJO E USO DO SOLO NO BRASIL	16
2.4	USO DA BIBLIOTECA DADA2 E PROCESSAMENTO 16S RRNA	17
2.5	USO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA ANÁLISE DE DADOS ASVS	18
2.6	TRABALHOS RELACIONADOS	19
3	MATERIAL E MÉTODOS	21
3.1	ÁREA EXPERIMENTAL E AMOSTRAGEM DO SOLO	21
3.2	DADOS DE ATRIBUTOS FÍSICOS, QUÍMICOS, BIOLÓGICOS E ÍNDICE DE SAÚDE DO SOLO	23
3.3	OBTENÇÃO DOS DADOS 16S RRNA	26
3.3.1	Biblioteca <i>Forward</i>	27
3.3.2	Análise Taxonômica	29
3.3.3	Normalização e Filtragem de Qualidade dos Dados DADA2	30
3.3.4	Ambiente Computacional	34
3.4	ALGORITMOS E BIBLIOTECAS ESTATÍSTICAS UTILIZADAS	34
3.4.1	Análise de Diversidade Alfa	34
3.4.2	Análise de Diversidade Beta	35
3.4.3	PCA e Partição de Variância	35
3.4.4	Algoritmo Random Forest	37
3.4.5	Algoritmo Extreme Gradient Boosting	37
3.4.6	Rede de Correlação de Spearman	38
3.4.7	Rede Espécies Indicadoras	39
4	RESULTADOS E DISCUSSÕES	41
4.1	ANÁLISE DE ABUNDÂNCIA RELATIVA DE FILOS PARA LOCAL E TRATAMENTO	41
4.2	ANÁLISE DE DIVERSIDADE ALFA DA COMUNIDADE BACTERIANA DOS TRATAMENTOS	42
4.3	ESTRUTURA DA COMUNIDADE MICROBIANA ATRAVÉS DA PCOA E NMDS COMPARADA A DBRDA	44
4.4	ANÁLISE PARTIÇÃO DE VARIAÇÃO	48
4.5	ESTRUTURAÇÃO DA COMUNIDADE BACTERIANA ATRAVÉS DE ALGORITMOS DE MACHINE LEARNING	51

4.6 ANÁLISE REDE DE CORRELAÇÃO SPEARMAN PARA USOS DO SOLO	56
4.6.1 Comparação entre redes - Nós Centrais.....	63
4.6.2 Comparação entre redes - Módulos	64
4.7 ANÁLISE REDE ESPÉCIE INDICADORA ATRAVÉS DO MÉTODO ÍNDICE DE VALOR INDICADOR COMPARADAS POR ÍNDICE LEVIN'S B	65
4.8 ANÁLISE ABUNDÂNCIA RELATIVA DE VIAS METABÓLICAS INFERIDAS POR PICRUST2.....	71
4.9 COMPARAÇÃO ENTRE ANÁLISES E TRATAMENTOS	74
4.9.1 Conservação da Floresta Secundária próximo ao uso agrícola do solo.....	79
5 CONCLUSÃO	81
REFERÊNCIAS	84

1 INTRODUÇÃO

O solo é um dos ecossistemas mais complexos e biologicamente diversos do planeta, abrigando uma vasta gama de microrganismos que desempenham papéis essenciais na ciclagem de nutrientes, na decomposição da matéria orgânica e nas interações com as plantas. Esses microrganismos, especialmente as bactérias, estão diretamente envolvidos na manutenção da fertilidade do solo, da produtividade agrícola e da sustentabilidade dos sistemas agroecológicos (Fierer, 2017).

A composição e a estrutura da comunidade bacteriana do solo são influenciadas por fatores ambientais, como pH, disponibilidade de nutrientes, umidade, uso do solo e tipo de manejo agrícola (Zinger et al., 2019; Etto et al., 2022). Mudanças no uso do solo, como a conversão de áreas de Floresta Secundária para pastagem ou para sistemas de cultivo, podem alterar significativamente a diversidade e a funcionalidade microbiana, comprometendo o equilíbrio ecológico do solo e sua capacidade de sustentar processos biogeoquímicos essenciais (Goss-Souza et al., 2019; Bobulska et al., 2021).

Com os avanços recentes em técnicas de sequenciamento de DNA e em ferramentas de bioinformática, tornou-se possível caracterizar a microbiota do solo em alta resolução taxonômica e funcional. Ferramentas como DADA2 e QIIME, combinadas a modelos estatísticos e algoritmos de aprendizado de máquina (AM), como *Random Forest* e XGBoost, permitem não apenas a identificação de microrganismos com precisão, mas também modelar como os atributos ambientais moldam a estrutura da comunidade microbiana (Cock et al., 2009; Zhu et al., 2024).

Neste contexto, este trabalho investiga o impacto de diferentes usos do solo (Floresta Secundária, Pastagem e sistemas agrícolas) sobre a estruturação da comunidade bacteriana em quatro localidades do Paraná. O foco principal é compreender as respostas ecológicas da microbiota frente ao manejo antrópico. Para isso, dados edáficos foram integrados ao sequenciamento RNAr 16S através de uma abordagem analítica complementar. O estudo emprega a Partição de Variação em conjunto com algoritmos de aprendizado de máquina como ferramentas para inferir a influência relativa dos atributos do solo sobre a composição microbiana, explorando os potenciais direcionadores que estruturam essas comunidades sob diferentes pressões ambientais.

Destarte, este trabalho tem como objetivo principal investigar como diferentes ecossistemas preservados e agrícolas do solo no estado do Paraná influenciam a composição e estrutura da comunidade bacteriana do solo ao avaliar o desempenho e a aplicabilidade de

algoritmos de aprendizado de máquina na modelagem da estrutura bacteriana, contrastando-os com métodos estatísticos tradicionais (Partição de Variação). Ao utilizar métricas de validação cruzada e análise de *Feature Importance*, o estudo busca não apenas identificar os atributos direcionadores da microbiota, mas validar metodologicamente como abordagens computacionais podem extrair padrões ecológicos de forma mais robusta e detalhada.

Os objetivos específicos são:

- Avaliar como os ecossistemas preservados e agrícolas do solo (Floresta Secundária, Pastagem e Agricultura) influenciam a composição da comunidade bacteriana do solo, com base em dados de rRNA 16S.
- Avaliar e comparar o poder preditivo e a robustez entre métodos estatísticos tradicionais (Partição de Variação) e algoritmos de aprendizado de máquina (*Random Forest*, *XGBoost* e *Gradient Boosting*), utilizando métricas de desempenho (R^2 de validação cruzada e RMSE) e a estabilidade da importância de variáveis (*Feature Importance*), aplicados a um conjunto de dados de microbioma do solo de alta dimensionalidade.
- Identificar os grupos taxonômicos mais sensíveis à mudança de uso do solo, propondo potenciais bioindicadores de saúde do solo.
- Explorar a complexidade das interações ecológicas entre atributos do solo e ASVs por meio de redes de correlação específicas por manejo.

2 REVISÃO DA LITERATURA

2.1 EVOLUÇÃO DO USO DO GENE 16S rRNA EM ANÁLISES MICROBIANAS

A utilização do gene 16S rRNA como marcador molecular representa um dos marcos mais significativos na biologia moderna. Historicamente, a classificação bacteriana era dependente de características fenotípicas e metabólicas observadas em cultura. No entanto, o trabalho pioneiro de Woese e Fox (1977) revolucionou essa visão ao propor o uso de sequências conservadas de rRNA para inferir relações evolutivas, resultando na redefinição da árvore das espécies e na descoberta do domínio Archaea. Esse gene foi estabelecido como um “cronômetro evolutivo” ideal por ser universalmente distribuído em procariotos, funcionalmente constante e não sujeito a altas taxas de transferência horizontal de genes (Woese, 1987).

A adoção do 16S rRNA impulsionou a chamada “revolução independente de cultivo”. Antes do uso molecular, estimava-se que menos de 1% das bactérias ambientais pudessem ser cultivadas em laboratório, um fenômeno conhecido como “*The Great Plate Count Anomaly*” (Staley; Konopka, 1985). O desenvolvimento da Reação em Cadeia da Polimerase (PCR) e o desenho de primers universais na década de 1990 permitiram, pela primeira vez, a detecção direta de microrganismos não cultiváveis a partir de DNA ambiental, revelando uma diversidade microbiana até então inexplorada (Amann et al., 1995; Pace, 1997).

Estruturalmente, a utilidade do gene 16S rRNA (aproximadamente 1.500 pares de base) reside na sua arquitetura única, composta por nove regiões hipervariáveis (V1 a V9) intercaladas por regiões altamente conservadas. Enquanto as regiões conservadas permitem o desenho de primers universais capazes de anelar em uma vasta gama de taxa, as regiões hipervariáveis acumulam mutações que permitem a diferenciação taxonômica entre gêneros e espécies (Riehle et al., 2012; Mizrahi-Man et al., 2013; Yarza et al., 2014; Gwak; Rho, 2020; Jin et al., 2024). A escolha da região variável a ser sequenciada influencia diretamente a cobertura taxonômica e a resolução da análise em diferentes ambientes (Bukin et al., 2019).

Atualmente, com o advento do sequenciamento de nova geração (NGS) e plataformas de alto rendimento, como Illumina MiSeq e NovaSeq, a análise do 16S rRNA atingiu uma nova escala. Diferente dos métodos iniciais de clonagem e sequenciamento Sanger, que analisavam poucas centenas de sequências, o NGS permite a geração de milhões de leituras simultâneas, possibilitando o estudo profundo de comunidades complexas, a identificação de taxa raros e a análise de diversidade beta e alfa em larga escala (Caporaso et al., 2011; Knight et al., 2018). Essas tecnologias consolidaram o gene 16S rRNA como o padrão-ouro na ecologia microbiana

contemporânea, devido à sua correlação com filogenia geral e capacidade de identificar padrões biogeográficos (Riehle et al., 2012; Mizrahi-Man et al., 2013; Gwak; Rho, 2020; Jin et al., 2024).

2.2 SAÚDE DO SOLO E INDICADORES DE QUALIDADE

O conceito de saúde do solo vai além da visão de fertilidade agrícola, sendo definido atualmente como a capacidade contínua do solo de funcionar como um sistema vivo, dentro dos limites do ecossistema e do uso da terra, para sustentar a produtividade biológica, manter a qualidade do ar e da água e promover a saúde de plantas, animais e (Doran; Zeiss, 2000; Lehmann et al., 2020). Essa abordagem de integração reconhece o solo não apenas como um substrato para crescimento vegetal, mas como um ecossistema complexo e dinâmico que fornece serviços ecossistêmicos essenciais, como a ciclagem de nutrientes, o sequestro de carbono e a regulação hídrica.

Para avaliar essa saúde, utilizam-se indicadores de qualidade do solo, que são propriedades mensuráveis sensíveis a mudanças no manejo e no clima. Esses indicadores são divididos em três categorias: físicos (ex: textura, densidade, porosidade, estabilidade de agregados), que determinam a estrutura e a aeração; químicos (ex: pH, capacidade de troca catiônica, teores de macro e micronutrientes, carbono orgânico), que governam a disponibilidade nutricional; e biológicos (ex: biomassa microbiana, atividade enzimática, respiração basal), que refletem a atividade metabólica do sistema (Kibblewhite et al., 2008; Bunemann et al., 2018).

Historicamente, as análises agrícolas focaram predominantemente nos indicadores químicos e físicos. No entanto, os indicadores biológicos ganharam destaque recente por sua alta sensibilidade e capacidade de atuar como “sinais de alerta precoce” (*early warnings*) para degradação ou recuperação do solo, respondendo a alterações de manejo muito antes que mudanças mensuráveis ocorram nas propriedades químicas ou físicas (Schloter et al., 2018). Nesse cenário, a integração de novos indicadores moleculares derivados do sequenciamento de DNA (como a diversidade e composição de ASVs) aos indicadores clássicos representa a nova fronteira para um diagnóstico mais preciso e funcional da saúde edáfica (Fierer et al., 2021; Yang et al., 2024).

2.3 SISTEMAS DE MANEJO E USO DO SOLO NO BRASIL

O Brasil consolidou-se nas últimas décadas como uma potência agrícola global, transitando de um modelo baseado na expansão de área para um paradigma de intensificação sustentável (EMBRAPA, 2023). Com a mudança do uso agrícola da terra, o modo como o solo é manejado exerce influência direta sobre suas propriedades físico-químicas e, consequentemente, sobre a estrutura e funcionalidade da comunidade microbiana (Szoboszlay et al., 2017; Arunrat et al., 2025; Caires et al., 2025). Os sistemas de produção variam desde monoculturas com intenso revolvimento do solo até sistemas integrados complexos, e tais sistemas integrados, por sua vez, têm sido adotados como estratégia sustentável para a produção de grãos como a soja, aumentando rendimento e resistência do solo (Soares et al., 2024; Souza et al., 2024; Vaupel et al., 2025). A soja, um dos grãos mais cultivados no país, muitas vezes está inserida nesses sistemas de produção variados, que equilibram alta produtividade e preservação do solo (EMBRAPA, 2023).

O Preparo Convencional (PC), caracteriza-se pelo revolvimento intensivo das camadas superficiais do solo por meio de aragem e gradagem. Embora facilite o estabelecimento inicial das culturas, essa prática rompe a estrutura dos agregados, expõe a matéria orgânica à oxidação rápida e destrói as hifas fúngicas, resultando frequentemente em perda de biodiversidade edáfica e erosão (Lal, 2015; Kaschuk et al., 2010). Em contrapartida, o Plantio Direto (PD), amplamente adotado no Brasil, fundamenta-se em três pilares: não revolvimento do solo, cobertura permanente (palhada) e rotação de culturas. O PD favorece a acumulação de carbono, a retenção de umidade e a preservação de nichos ecológicos para a microbiota (EMBRAPA, 2009; EMBRAPA, 2025).

As Pastagens ocupam a maior parte da área agricultável do país. No entanto, grande parte encontra-se em algum estágio de degradação, o que compromete a ciclagem de nutrientes. O manejo adequado de pastagens, com reposição de fertilidade e ajuste de carga animal, é importante para a manutenção da atividade biológica do solo (EMBRAPA, 2014). Como evolução desses conceitos, surgem os Sistemas Silvopastoris (SS) e a Integração Lavoura-Pecuária-Floresta (ILPF). Esses sistemas mimetizam processos naturais ao consorciar árvores, forrageiras e animais na mesma área. A presença do componente arbóreo modifica o microclima, diversifica o aporte de serapilheira e cria um sistema radicular estratificado, promovendo uma maior diversidade microbiana e resiliência ecossistêmica quando comparados a monoculturas ou pastagens (EMBRAPA., 2011; Souza et al., 2024; Vaupel et al., 2025).

2.4 USO DA BIBLIOTECA DADA2 E PROCESSAMENTO 16S rRNA

Antes da consolidação dos métodos baseados em Variantes de Sequência de Amplicon (ASVs), a análise de sequenciamento 16S rRNA era dominada pelo uso de Unidades Taxonômicas Operacionais (OTUs), agrupadas tipicamente por similaridade de 97% (Edgar, 2013; Schloss et al., 2009). Embora amplamente utilizadas historicamente, as OTUs apresentavam limitações analíticas importantes: a dependência de limiares de similaridade fixos e arbitrários, o agrupamento de sequências biologicamente distintas dentro de um mesmo módulo e a dificuldade de comparação entre estudos devido à falta de padronização nos algoritmos de *clustering* (Callahan et al., 2017; Nearing et al., 2018). Nesse contexto, o pipeline DADA2 (Callahan et al., 2016) promoveu um avanço significativo ao introduzir um modelo explícito de erros de sequenciamento, capaz de distinguir erros de amplificação de variantes biológicas reais, gerando ASVs com resolução de nucleotídeo único. Essa abordagem reduziu a inflação artificial de diversidade comum em métodos baseados em OTUs e aumentou a sensibilidade para detectar variações biológicas sutis (Prodan et al., 2020). Além disso, por representarem sequências exatas, as ASVs permitem a reprodutibilidade e a comparação direta entre diferentes estudos e plataformas, funcionando como um código de barras biológico consistente (Glockner et al., 2017; Knight et al., 2018). A implementação de filtragem de qualidade orientada por modelos e a remoção de quimeras consolidaram o DADA2 como uma ferramenta robusta e padronizada para estudos de microbioma (Callahan et al., 2016).

Em sistemas de manejo orgânico versus convencional, através do pipeline do DADA2, os dados gerados, permitiram identificar diferenças relevantes na riqueza bacteriana, composição e estrutura de rede entre práticas de cultivo (Armalyte et al., 2019; Maretto et al., 2023; Lori et al., 2024). Dados de 16S rRNA processados também se mostraram sensíveis o suficiente para detectar efeitos de diferentes estratégias de plantio na diversidade e abundância de microrganismos do solo (Narayana et al., 2022; Rokins et al., 2022). Os dados gerados podem ser aplicados em análises estatísticas, tais como, PCA (análise de componentes principais), análise de abundância relativa e PCoA (análise de coordenadas principais), trazendo informações importantes sobre a formação comunidade microbiana do solo, em diferentes usos do solo (Wang et al., 2020; Rijssel et al., 2022; Maretto et al., 2023; Liu J. et al., 2024).

2.5 USO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA ANÁLISE DE DADOS ASVS

A interpretação de dados ecológicos e microbiológicos baseou-se predominantemente em métodos de estatística multivariada clássica. Abordagens baseadas em ordenação, como a Análise de Componentes Principais (PCA), Análise de Coordenadas Principais (PCoA) e o Escalonamento Multidimensional Não-Métrico (NMDS), combinadas a testes de permutação como PERMANOVA (Análise de Variância Multivariada Permutacional), detalhavam a estrutura de comunidades (Ramette, 2007; Anderson, 2017). Essas ferramentas permitem visualizar dissimilaridades entre grupos e testar hipóteses gerais sobre a variância beta-diversidade, permitindo inferir como tratamentos ou locais diferem globalmente.

Contudo, com o avanço do sequenciamento de alto desempenho e a geração de grandes volumes de dados taxonômicos (como os baseados no gene 16S rRNA), as limitações das abordagens clássicas tornaram-se evidentes. Os dados de microbioma, agora frequentemente analisados no nível de ASVs (Amplicon Sequence Variants), apresentam alta dimensionalidade esparsidade excessiva e relações complexas e não lineares que métodos lineares tradicionais falham em capturar adequadamente (Knights et al., 2011; Dudek et al., 2024). A necessidade de ir além da simples descrição das diferenças entre comunidades para a predição de características e a seleção robusta de biomarcadores impulsionou a adoção de novas metodologias computacionais.

Os algoritmos de AM consolidaram-se como ferramentas necessárias, complementando as análises estatísticas clássicas. Diferente dos testes tradicionais, o AM é capaz de lidar eficientemente com a alta dimensionalidade e a complexidade ecológica inerente aos dados de ASVs. Algoritmos baseados em árvores de decisão, como *Random Forest* (RF) e XGBoost, destacam-se por sua eficácia em relacionar a composição microbiana com atributos do solo, como pH, matéria orgânica, textura e indicadores biológicos.

Essas técnicas permitem não apenas identificar quais ASVs ou filos são mais sensíveis a alterações no uso do solo, mas também prever atributos do solo e o quanto a comunidade bacteriana influencia na sua estruturação, com base em dados taxonômicos com acurácia, superando limitações de modelos lineares simples. Em solos agrícolas, por exemplo, modelos de RF e XGBoost têm sido aplicados para destacar variáveis-chave que explicam a estrutura da comunidade e na construção de bioindicadores microbiológicos úteis em programas de monitoramento ambiental e avaliação da saúde do solo (Hermans et al., 2020; Wilhelm et al.,

2022; Lei et al., 2023; Knight et al., 2024; Lei et al., 2024; Roopashree et al., 2024; Alonso-Sarria et al., 2025).

2.6 TRABALHOS RELACIONADOS

A conversão de ecossistemas nativos em sistemas agrícolas é reconhecida como um dos principais fatores de alteração da estrutura e função das comunidades microbianas do solo em escala global. Hermans et al. (2020) e Maretto et al. (2023) trabalharam com análises em diferentes tratamentos de solo, usando técnicas estatísticas para cada atributo a ser analisado, como por exemplo o uso da PCA para agrupar amostras de solo agrícola com base em características químicas e físicas, e a PCoA para analisar a diversidade beta da comunidade de bactérias do solo. Outros trabalhos são apresentados no Quadro 1.

As análises utilizadas por diferentes autores e em diferentes locais e usos do solo, são similares e podem contribuir para extrair informações importantes do solo. O uso de AM tem se demonstrado um complemento de alta precisão, através dos dados gerados pela análise do 16S rRNA de organismos microbianos (ou fúngicos), reforçando a estatística multivariada.

Quadro 1 - Trabalhos relacionados a análises estatísticas a partir dos dados de 16s de microrganismos do solo
(continua)

AUTOR	ANÁLISES UTILIZADAS	USO DO SOLO
Hermans et al. (2020)	Diferenças composicionais relativas (dissimilaridade de Bray-Curtis), PCA (com comparativo <i>Random Forest</i>), <i>Clusters</i> de atributos químicos do solo usando <i>Random Forest</i>	Área Indígena, Floresta Exótica e Área Pastagem
Wang et al. (2020)	Índice de riqueza/Shannon (diversidade alfa)	Plantio Orgânico
Fan et al. (2022)	Índice de riqueza por CHAO1, ACE e Shannon (diversidade alfa), PCoA (diversidade beta), PCA, Rede de co-ocorrência	Área Floresta, Casa de vegetação, Plantio Direto (árvores de chá)
Rijssel et al. (2022)	Índice de riqueza/Shannon, PCoA (diversidade beta)	Plantio Orgânico e Plantio Convencional

Quadro 1 - Trabalhos relacionados a análises estatísticas a partir dos dados de 16s de microrganismos do solo

(conclusão)

AUTOR	ANÁLISES UTILIZADAS	USO DO SOLO
Wilhelm et al. (2022)	Índice de riqueza/Shannon/Pielou, PERMANOVA e covariância, SVM e <i>Random Forest</i>	Plantio Convencional
Lei et al. (2023)	Índice de riqueza/Shannon/Pielou/Chao1, PERMANOVA e PCoA (diversidade beta), Correlação de Pearson, Rede de co-ocorrência	Diversos
Maretto et al. (2023)	PCA, PCoA (diversidade beta), Índice de Riqueza/Shannon/Simpson, Abundância Relativa	Plantio Orgânico e Plantio Convencional
Guo et al. (2023)	Índice de riqueza/CHAO1/Shannon/Pielou, PCoA, PICRUST2, Teste Parcial de Mantel, Partição de Variância, Teste de permutação Monte Carlo e <i>Random Forest</i> , Correlação de Pearson	Florestas de pinheiros naturais
Knight et al. (2024)	NMDS, RDA, Dissimilaridade de Bray-Curtis, PERMANOVA (diversidade beta), <i>Random Forest</i> , Correlação de Pearson	Pastagem
Liu J. et al. (2024)	Índice de riqueza/Simpson/Pielou, Abundância Relativa, Correlação de Pearson, Análise de Redundância (RDA)	Plantio Convencional, <i>Intercropping</i> (Consórcio)

Fonte: O autor.

3 MATERIAL E MÉTODOS

Este estudo visa investigar como diferentes ecossistemas preservados e agrícolas do solo no estado do Paraná influenciam a composição e estrutura da comunidade bacteriana do solo, com base em atributos físicos, químicos e biológicos, através de técnicas de bioinformática e AM.

Nesta seção, serão apresentados como os dados foram utilizados e como foi realizado as análises utilizando as técnicas de bioinformática.

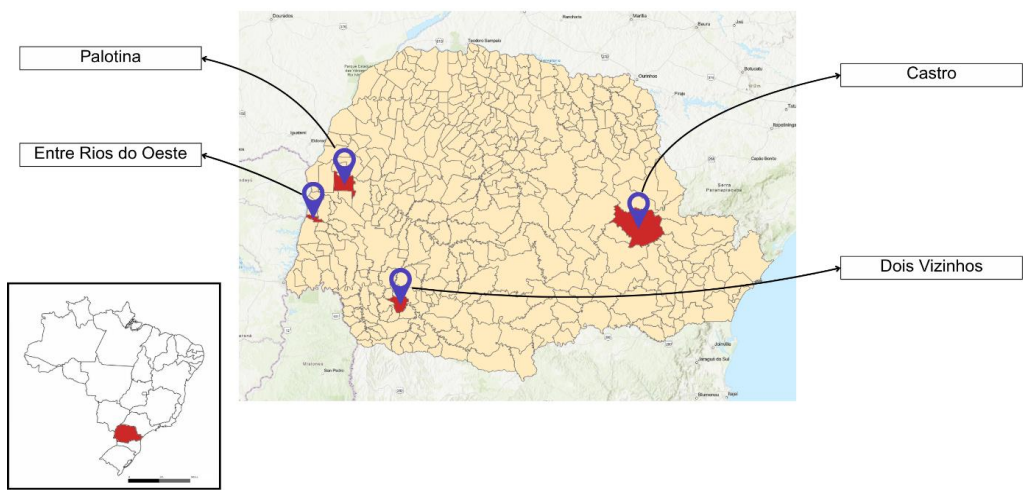
3.1 ÁREA EXPERIMENTAL E AMOSTRAGEM DO SOLO

Para a composição do conjunto de dados deste estudo, foram selecionadas quatro localidades distribuídas em mesorregiões distintas do estado do Paraná, Sul do Brasil (Figura 1): Castro, Dois Vizinhos, Entre Rios e Palotina. Foram avaliados, no total, cinco sistemas de uso do solo, abrangendo áreas de referência preservadas (Floresta Secundária) e ecossistemas agropecuários (Plantio Direto, Preparo Convencional, Pastagem e Sistema Silvipastoril). O delineamento das amostras não foi simétrico entre as regiões: enquanto a Floresta Secundária e o Plantio Direto foram amostrados em todas as quatro localidades, o terceiro sistema de manejo avaliado variou de acordo com a prática regional. Especificamente, avaliou-se a Pastagem em Castro e Entre Rios; o Sistema Silvipastoril em Dois Vizinhos; e o Preparo Convencional em Palotina. Quatro locais foram amostrados entre Agosto e Outubro no ano de 2020 (Quadro 2), em cada local foi coletado solo de 3 manejos diferentes. Para cada uso do solo (manejo) foi implantado uma malha regular de 9 X 9 metros, e coletadas 9 amostras de solo. No total foram obtidos, portanto, 108 amostras de solo (4 locais x 3 diferentes manejos x 9 repetições -

Quadro 3). A amostragem de solo foi realizada na camada de 0-10 cm para os indicadores biológicos, físicos e químicos do solo. Os solos coletados para as avaliações biológicas foram acondicionados em caixas térmicas com gelo e enviados ao Laboratório (Becker et al., 2025).

Os locais selecionados para estudo, são as amostras coletadas nos municípios de Castro, Dois Vizinhos, Entre Rios do Oeste e Palotina (sendo estes representantes de diferentes altitudes e distribuição de atributos físicos, químicos e biológicos distintos).

Figura 1 - Locais de estudo no estado do Paraná



Fonte: O autor.

Quadro 2 - Descrição dos locais de estudo

LOCAIS DE ESTUDO	MUNICÍPIO	COORDENADAS GEOGRÁFICAS	ALTITUDE (M)	REGIÃO EDAFOCLI-MÁTICA ²	CLIMA ³
LBw-CAS	Castro	24°49'26,3"S 49°58'01,0"W	1025	103	Cfb
LVdf-DOV	Dois Vizinhos	25°41'19,5"S, 53°05'29,7"W	504	102	Cfa
LVef-ERO	Entre Rios do Oeste	24°40'32,54"S, 54°16'52,79"W	243	102	Cfa
LVef-PAL	Palotina	24°20'44,92"S, 53°45'17,03"W	363	201	Cfa

¹ Classificação do solo segundo EMBRAPA (2007). ² Região edafoclimática é caracterizada por suas condições de solo (proporção de argila, areia e silte) e clima que influenciam o desenvolvimento de plantas, afetando principalmente a produção de grãos (EMBRAPA, 2015). ³ Clima Cfb: clima subtropical sem estação seca e verão fresco (temperatura média do mês mais quente inferior a 22°C). Clima Cfa: clima subtropical sem estação seca e verão quente (temperatura média do mês mais quente superior a 22°C), de acordo com a classificação de Koppen (Alvares et al., 2013). Fonte: Adaptado de Becker et al., 2025.

Quadro 3 - Composição do conjunto de dados segundo a localidade de amostragem, município e as classes de uso e manejo do solo

LOCAIS DE ESTUDO	MUNICÍPIO	USO E MANEJO DO SOLO
LBw-CAS	Castro	Floresta Secundária (FS)
		Plantio Direto (PD)
		Pastagem (P)
LVdf-DOV	Dois Vizinhos	Floresta Secundária (FS)
		Plantio Direto (PD)
		Sistema Silvipastoril (SS)
LVef-ERO	Entre Rios do Oeste	Floresta Secundária (FS)
		Plantio Direto (PD)
		Pastagem (P)
LVef-PAL	Palotina	Floresta Secundária (FS)
		Plantio Direto (PD)
		Preparo Convencional (PC)

Fonte: O autor.

3.2 DADOS DE ATRIBUTOS FÍSICOS, QUÍMICOS, BIOLÓGICOS E ÍNDICE DE SAÚDE DO SOLO

Para uma caracterização abrangente do solo, foram coletadas amostras tanto perturbadas quanto não perturbadas. Amostras de solo não perturbadas para análises físicas foram obtidas com um amostrador Uhland com anéis de aço de 0,05 m de diâmetro e 0,05 m de altura. Amostras perturbadas para análises biológicas e químicas foram coletadas com um trado de aço inoxidável (5 cm de diâmetro), após a remoção da cobertura superficial. Além disso, blocos de solo semi-intactos (10 × 10 × 10 cm) foram coletados com uma pá reta para análise da estabilidade dos macroagregados.

As análises biológicas foram realizadas no Laboratório de Biologia Molecular Microbiana (LABMOM) da UEPG (Becker et al., 2025). A atividade da β -glicosidase

(atividade BG) foi determinada pelo método do p-nitrofenil- β -D-glicosídeo (Eivazi e Tabatabai, 1988; Tabatabai et al., 1982). A atividade da arilsulfatase foi determinada conforme Tabatabai & Bremner (1970). A atividade da fosfatase ácida foi medida utilizando p-nitrofenil fosfato como substrato (Tabatabai, 1994) com adaptações propostas por Balota et al. (2013), com base na determinação do p-nitrofenol (PNP) liberado após a incubação do solo com 0,05 M de 4-nitrofenil- β -glicósido a 410 nm. A atividade da urease foi determinada pela quantificação da liberação de amônia a partir da hidrólise da ureia (Kandeler e Gerber, 1988).

O carbono da biomassa microbiana (CMB) foi quantificado por meio do método de fumigação-extração (Vance et al., 1987). A respiração basal do solo (RBS) foi determinada após 7 dias de incubação em condições ótimas (25°C, 60% de espaço de poros preenchido por água), utilizando o método de Stotzky (1965). O quociente metabólico ($q\text{CO}_2$) foi calculado como a razão entre RBS e CMB ($\mu\text{g CO}_2\text{-C g}^{-1}\text{ CMB-C h}^{-1}$), fornecendo um índice da eficiência microbiana na utilização do carbono.

As proteínas do solo relacionadas à glomalina foram extraídas conforme Wright & Upadhyaya (1998). A proteína total relacionada à glomalina do solo (T-GRSP) foi extraída utilizando citrato de sódio 50 mM (pH 8,0) a 121°C, enquanto a proteína relacionada à glomalina facilmente extraível (EE-GRSP) foi obtida utilizando citrato de sódio 20 mM a 95°C. As concentrações de proteína foram determinadas pelo ensaio de Bradford, com albumina sérica bovina como padrão (Bradford, 1976).

A análise química do solo seguiu protocolos padronizados realizados no laboratório de Nutrição de Plantas da UEPG (Silva, 2009). Para determinação do pH em cloreto de cálcio (CaCl_2), utilizou-se uma relação solo: solução de 1:2,5 (solução 0,01 mol L^{-1}). Após a medição do pH em CaCl_2 , foram adicionados 4 mL de solução tampão SMP para a determinação da acidez potencial ($\text{H}^+ + \text{Al}^{3+}$). Os teores de potássio trocável (K) e fósforo disponível (P) foram extraídos com solução Mehlich-1, em uma relação solo: solução de 1:10. Para a determinação de K, as leituras foram realizadas em fotômetro de chama. Para P, a leitura foi realizada por espectrofotometria de absorção molecular em modo de transmitância a 660 nm. A extração de alumínio (Al), cálcio (Ca) e magnésio (Mg) foi realizada em solução de cloreto de potássio (KCl) 1 mol L^{-1} (relação solo: solução de 1:12,5). A concentração de Al foi determinada por titulação com NaOH, na presença de azul de bromotimol como indicador. As concentrações de Ca e Mg foram determinadas por complexometria com ácido etilenodiamino tetra-acético (EDTA). A capacidade de troca catiônica (CTC) e a saturação por bases (V%) foram calculadas seguindo procedimentos padrão. A CTC foi estimada como a soma das bases trocáveis (Ca^{2+} , Mg^{2+} e K^+) e da acidez potencial ($\text{H}^+ + \text{Al}^{3+}$). A V% foi calculada como a proporção das bases

trocáveis em relação à CTC. O carbono orgânico total (COT) foi determinado por oxidação úmida, segundo o método de Walkley-Black (Cantarella et al., 2001), com fatores de correção para oxidação incompleta.

Os indicadores físicos foram analisados no laboratório de Física do Solo da UEPG (Becker et al., 2025). Os teores de areia, silte e argila em cada local foram determinados pelo método do hidrômetro (Gee e Or, 2002). As amostras de solo não perturbadas foram saturadas por capilaridade e, em seguida, utilizadas para determinação da macroporosidade (MaP) sob tensão de -6 kPa e capacidade de campo (CC) sob tensão de -10 kPa em estação de laboratório pF (ecoTec). Após secagem em estufa a 105 °C, foram determinadas a microporosidade (MiP), a porosidade total (PT) e a densidade do solo (BD) (Teixeira et al., 2017). O ponto de murcha permanente (PMP) das plantas (-1500 kPa) foi calculado conforme Campbell (2006). A água disponível para as plantas (ADP) foi calculada como a diferença entre CC e PMP. Além disso, foi calculado o espaço de poros preenchido por água (WFPS) conforme Reynolds et al. (2002), como a razão entre CC e PT.

A estabilidade de agregados foi avaliada utilizando a metodologia de peneiramento úmido (Elliott, 1986) para determinar a distribuição das classes de agregados na estabilidade de macroagregados (AGS) (>250 µm). Este parâmetro fornece informações sobre a estabilidade estrutural do solo e sua relação com a atividade biológica. A resistência à penetração do solo (SRP) foi medida em amostras não perturbadas sob potencial matricial de -10 kPa, utilizando um penetrômetro eletrônico estático de laboratório, com leituras realizadas no centro geométrico de cada amostra (Tormena et al., 1998).

Os dados dos atributos físicos que se referem a quantidade de argila, silte, areia, densidade, macroporosidade, microporosidade, porosidade total, capacidade de campo, ponto de murcha permanente, água disponível, resistência a penetração, macroagregados, microagregados presentes no solo e a classe textural do solo dos locais e manejos analisados. A determinação da capacidade de troca de cátions (CTC), relação cálcio e magnésio (Ca:Mg), percentagem de saturação por bases (V%), pH (CaCl₂ e água), alumínio (Al), cálcio (Ca), magnésio (Mg), fósforo (P), potássio (K). Os dados biológicos utilizados se referem às atividades das enzimas relacionadas com o ciclo biogeoquímico do carbono, enxofre, nitrogênio e fosfato, denominadas β-glicosidase, Arilsulfatase, Urease, carbono orgânico total (COT), e Fosfatase ácida, respectivamente. Também a quantificação do carbono da biomassa microbiana (CBM), quociente metabólico do solo (qCO₂), quociente microbiano (qMIC), a respiração basal, glomalina total e glomalina facilmente extraível.

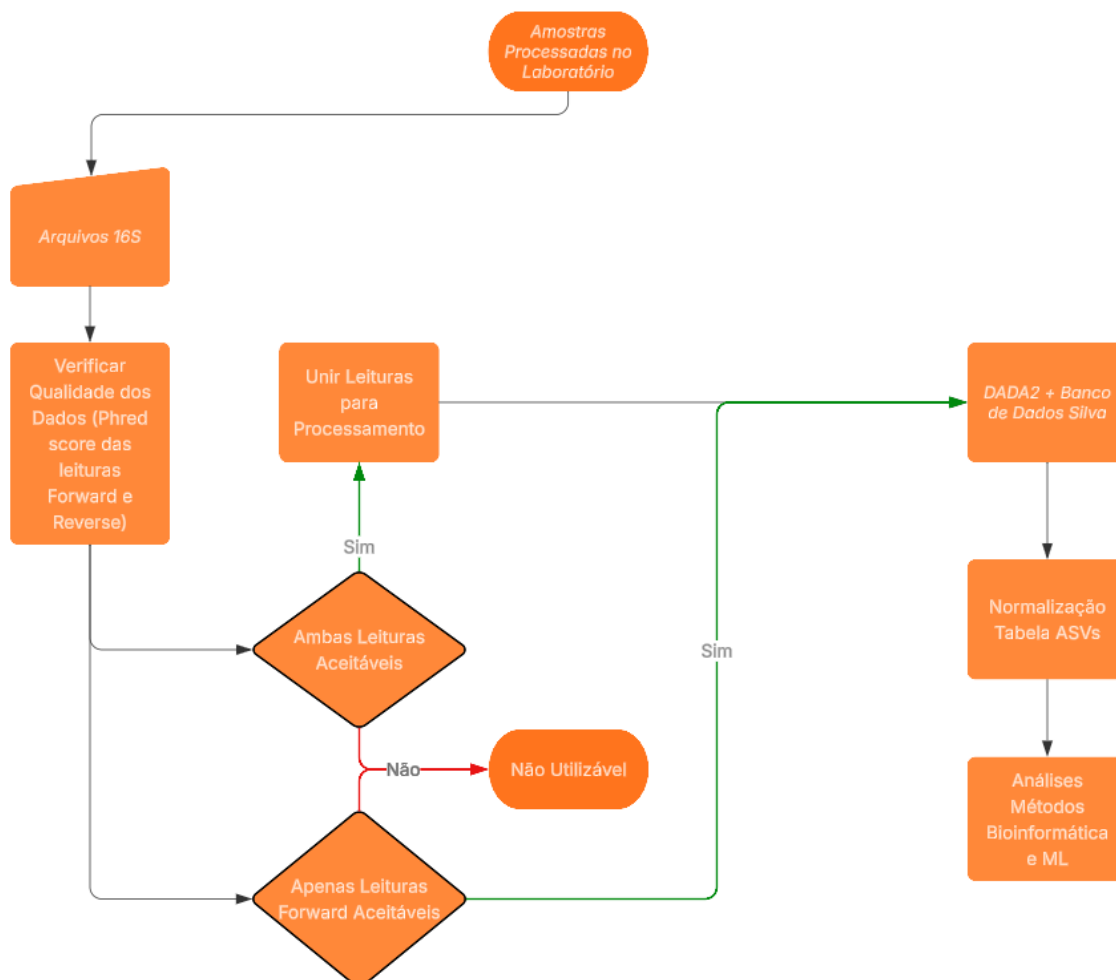
3.3 OBTENÇÃO DOS DADOS 16S RRNA

Para a caracterização da comunidade bacteriana presente nas amostras de solo, foi realizada a extração do DNA total diretamente do solo, utilizando-se um protocolo específico para amostras ambientais, com o objetivo de garantir a recuperação de material genético representativo e de alta qualidade. Após a extração, a região hipervariável V3-V4 do gene 16S rRNA foi amplificada por PCR (amplificação de uma ou mais cópias de um segmento de DNA em várias ordens de grandeza, gerando entre milhares a milhões de cópias de uma determinada sequência de DNA), utilizando primers universais com adaptadores compatíveis com a plataforma de sequenciamento *Illumina* (Sinclair et al., 2015, Pichler et al., 2018).

Os produtos amplificados foram quantificados, purificados e posteriormente agrupados em uma biblioteca de sequenciamento. O sequenciamento foi realizado com a tecnologia *Illumina MiSeq*, com a geração de leituras pareadas (paired-end), cerca de 440 pares de base, os *primers* universais utilizados para o alinhamento são 515F-Y e 926R. A PCR foi realizada em triplicata para cada amostra com: 94°C por 3 min; 18 ciclos de 94°C por 45 s, 50°C por 30 s e 68°C por 60 s; seguida por 72°C por 10 min. As sequências obtidas foram então submetidas a uma etapa de controle de qualidade, remoção de quimeras e agrupamento em unidades taxonômicas operacionais (OTUs) ou amplicon sequence variants (ASVs), conforme o pipeline analítico adotado (DADA2).

Essa abordagem possibilitou a identificação e a análise da diversidade microbiana presente nas amostras de solo, com base no perfil taxonômico derivado das sequências do gene 16S rRNA, seguindo o fluxo de processamento proposto, como exemplificado na Figura 2, onde figura representa o fluxograma metodológico do pipeline de processamento dos dados de sequenciamento do gene 16S rRNA. O processo inicia com a verificação de qualidade (score Phred), seguida pela decisão estratégica de união das leituras (paired-end merge) ou utilização exclusiva das leituras forward dependendo da qualidade. A inferência de Variantes de Sequência de Amplicon (ASVs) e a classificação taxonômica foram realizadas via DADA2 utilizando o banco de dados SILVA, culminando na normalização da tabela para as análises estatísticas e de AM.

Figura 2 - Fluxograma Método Processamento 16S rRNA



Fonte: O autor.

3.3.1 Biblioteca *Forward*

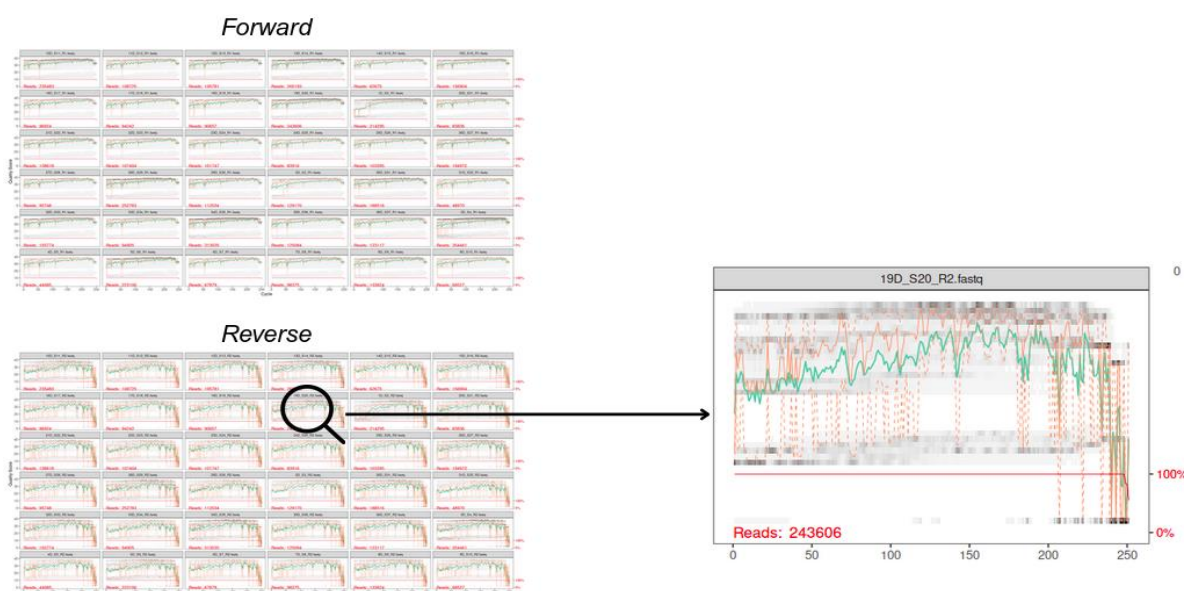
Durante a análise inicial dos perfis de qualidade (*Phred score*) das leituras geradas pela plataforma *Illumina*, observou-se que as sequências *reverse* apresentavam qualidade significativamente inferior em comparação às *forward*, especialmente em suas regiões iniciais e finais, onde identifica-se na Figura 3, a figura representa perfis de qualidade das leituras de sequenciamento forward (painel superior) e reverse (painel inferior) avaliadas pelo score Phred. O painel ampliado à direita, representativo da amostra 19D_S20_R2, ilustra a queda acentuada na qualidade e o aumento do ruído (linhas tracejadas vermelhas) nas bases finais das leituras reverse. Com base neste padrão inconsistente, observável na maioria das amostras, optou-se por utilizar exclusivamente as leituras forward para as etapas subsequentes de análise. Esse padrão

é recorrente em dados de sequenciamento *Illumina*, já que os reagentes e a intensidade do sinal decaem ao longo das múltiplas rodadas de leitura, resultando em maior erro nas *reverse* (Guo et al., 2013; Tan et al., 2019).

Foi adotada a estratégia de *single-end* para lidar com sequências *reverse* de baixa qualidade, incluindo a opção de descartar essas leituras e manter apenas as *forward*, quando a qualidade da segunda leitura prejudica a montagem ou análise posterior (Werner et al., 2011; Kwon et al., 2013). Entre as vantagens dessa abordagem está a redução da perda de dados durante a filtragem, além de manutenção de cobertura suficiente para inferência de diversidade e abundância microbiana sem introduzir ruído adicional.

A biblioteca escolhida, DADA2 (Callahan et al., 2016), oferece suporte para análise em modo *single-end*, permitindo o processamento eficiente das sequências *forward* isoladamente (“*denoise-single*”), preservando a robustez taxonômica dos resultados mesmo sem dependência das leituras *reverse* (Wu et al., 2015; Abellan-Schneyder et al., 2021; Silva et al., 2023).

Figura 3 - Análise de qualidade biblioteca de sequências RNA



Fonte: O autor.

3.3.2 Análise Taxonômica

A análise taxonômica das sequências do gene 16S rRNA foi conduzida utilizando a linguagem R, com as bibliotecas “DADA2” e “Biostrings”, em modo *single-end*, considerando apenas as sequências *forward* devido à baixa qualidade das *reverse*. O *pipeline* seguiu as etapas recomendadas na documentação oficial do “DADA2”, com as seguintes etapas a seguir.

As sequências *forward* foram processadas por meio da função “*filterAndTrim()*”, com o parâmetro “*truncLen=210*”, visando remover regiões de baixa qualidade e manter uma base uniforme para análise subsequente. Essa etapa também inclui o descarte de sequências com “N”, baixa qualidade média ou presença de “*phiX*” (bacteriófago que é utilizado como controle de qualidade em experimentos de sequenciamento da plataforma *Illumina*). As sequências filtradas foram submetidas à função “*learnErrors()*”, que estima a taxa de erro específica dos dados a partir de um subconjunto de amostras, permitindo uma correção precisa dos erros durante o processo de inferência.

A função “*dada()*” foi aplicada às leituras filtradas, utilizando o modelo de erro previamente aprendido. Essa etapa implementa o algoritmo DADA (*Divisive Amplicon Denoising Algorithm*), que infere as variantes reais presentes nas amostras, corrigindo os erros de sequenciamento e evitando a formação de OTUs arbitrárias. As ASVs inferidas foram organizadas em uma matriz de abundância por meio da função “*makeSequenceTable()*”, que consolida as sequências únicas e suas respectivas frequências em cada amostra. A função “*removeBimeraDenovo()*” foi utilizada para remover sequências quiméricas de forma de novo, baseando-se nas abundâncias relativas e na semelhança com outras ASVs da própria amostra.

As sequências não quiméricas foram extraídas utilizando a função “*getSequences()*” e convertidas em um objeto do tipo “*DNAStringSet()*” (da biblioteca “*Biostrings*”), etapa necessária para a atribuição taxonômica subsequente. A classificação taxonômica foi realizada utilizando a função “*IdTaxa()*”, com base no banco de dados SILVA (SSU r138.2 – 2024) (Quast et al., 2013), versão apropriada para análise do gene 16S rRNA. O método emprega um classificador baseado em k-mer treinado a partir da base de referência, fornecendo identificação taxonômica hierárquica até o nível de gênero (ou espécie, quando possível). Identificado cada amostra segundo o Quadro 4, de acordo com localidade, numeração de réplica e tratamento.

Quadro 4 – Representação dos nomes de pontos de coleta

CIDADE	RÉPLICA	TRATAMENTO	NOME DA AMOSTRA
Castro	1,2,3	Floresta Secundária	C1FS / C2FS / C3FS
Castro	1,2,3	Pastagem	C1PA / C2PA / C3PA
Castro	1,2,3	Plantio Direto	C1PD/ C2PD / C3PD
Dois Vizinhos	1,2,3	Floresta Secundária	DV1FS / DV2FS / DV3FS
Dois Vizinhos	1,2,3	Pastagem	DV1PA / DV2PA / DV3PA
Dois Vizinhos	1,2,3	Plantio Direto	DV1PD / DV2PD / DV3PD
Entre Rios	1,2,3	Floresta Secundária	Er1FS / Er2FS / Er3FS
Entre Rios	1,2,3	Sistema Silvopastoril	Er1SS / Er2SS / Er3SS
Entre Rios	1,2,3	Plantio Direto	Er1PD / Er2PD / Er3PD
Palotina	1,2,3	Floresta Secundária	Pa1FS / Pa2FS / Pa3FS
Palotina	1,2,3	Preparo Convencional	Pa1PC / Pa2PC / Pa3PC
Palotina	1,2,3	Plantio Direto	Pa1PD / Pa2PD / Pa3PD

Fonte: O autor.

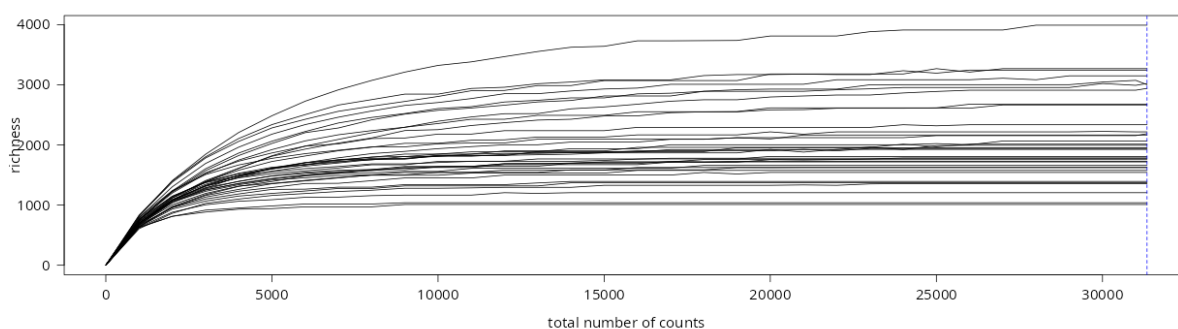
3.3.3 Normalização e Filtragem de Qualidade dos Dados DADA2

Após a obtenção da tabela de abundância de ASVs, foi realizada a normalização dos dados para compensar variações no número de sequências entre amostras, garantindo comparabilidade nas análises de diversidade (Weiss et al., 2017). Esse procedimento é crucial, pois diferenças no esforço de sequenciamento podem distorcer estimativas de riqueza e composição microbiana, influenciando análises de diversidade alfa e beta.

Optou-se pelo método SRS (*“Scaling with Ranked Subsampling”*, escalonamento com subamostragem classificada) (Beule; Karlovsky, 2020), presente no pacote “SRS”, que preserva a estrutura relativa da comunidade ao normalizar amostras para um número fixo de leituras (C_{min}), definido conforme o platô identificado, Figura 4. Neste processo foi realizada a avaliação da profundidade mínima através da curva de rarefação. A curva resultante mostrou um platô a 32.082 leituras por amostra, indicando que acima desse valor a adição de novas leituras não geraria aumento significativo na riqueza. Esse ponto serve como corte lógico para a normalização por rarefação controlada.

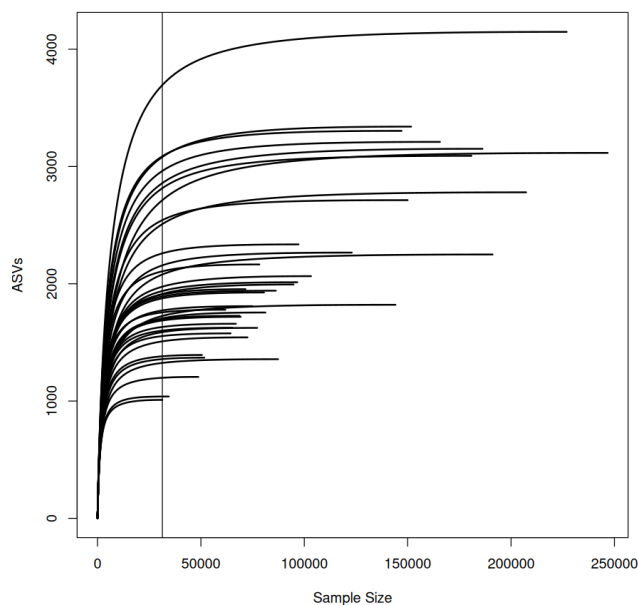
O pacote aplica escalonamento proporcional seguido de *subsampling* ranqueado, minimizando erros de amostragem e garantindo consistência na distribuição relativa das ASVs. Como resultado, foi gerada uma nova tabela de abundância, padronizada em 32.082 leituras/amostra (Figura 7). Na Figura 7, é analisado graficamente a tabela de dados de contagem de ASVs, obtido através do processo DADA2 e base de dados SILVA (SSU r138.2 – 2024), ambos os gráficos exibem a abundância relativa de ASVs para cada local de coleta, sendo o primeiro gráfico (à esquerda) antes do processo de normalização e corte utilizando método SRS (Figura 5) e o segundo gráfico (à direita) após processo realizado, cada barra demonstra a abundância para cada domínio presente na biblioteca, sendo o domínio *Bacteria* (cor verde) o mais abundante.

Figura 4 - Curvas SRS para cada Tratamento



Fonte: O autor

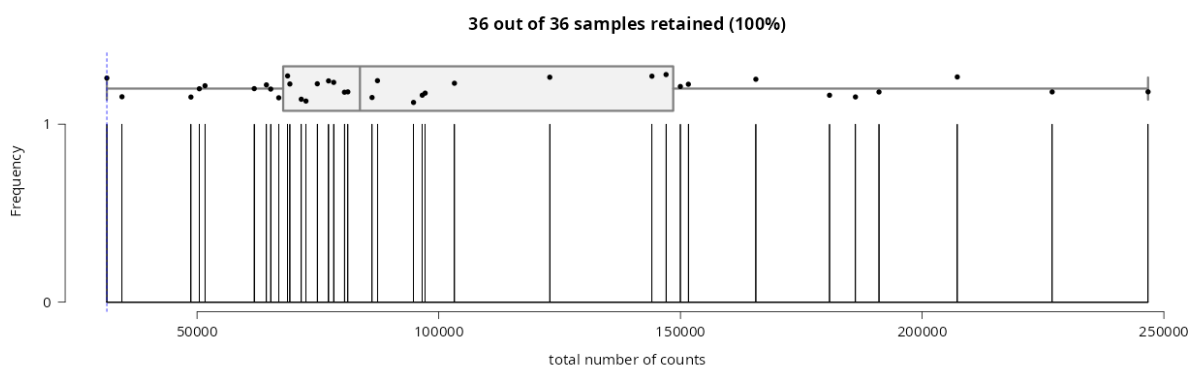
Figura 5 - Corte amostras através do SRS



Fonte: O autor

O processo de filtrar utilizando o método SRS, manteve a preservação das 36 amostras dos tratamentos, ou seja, a biblioteca identificou que todas as amostras são válidas e não possuem inconsistências, como identificado na Figura 6.

Figura 6 - Frequência de ASVs nas Amostras por SRS



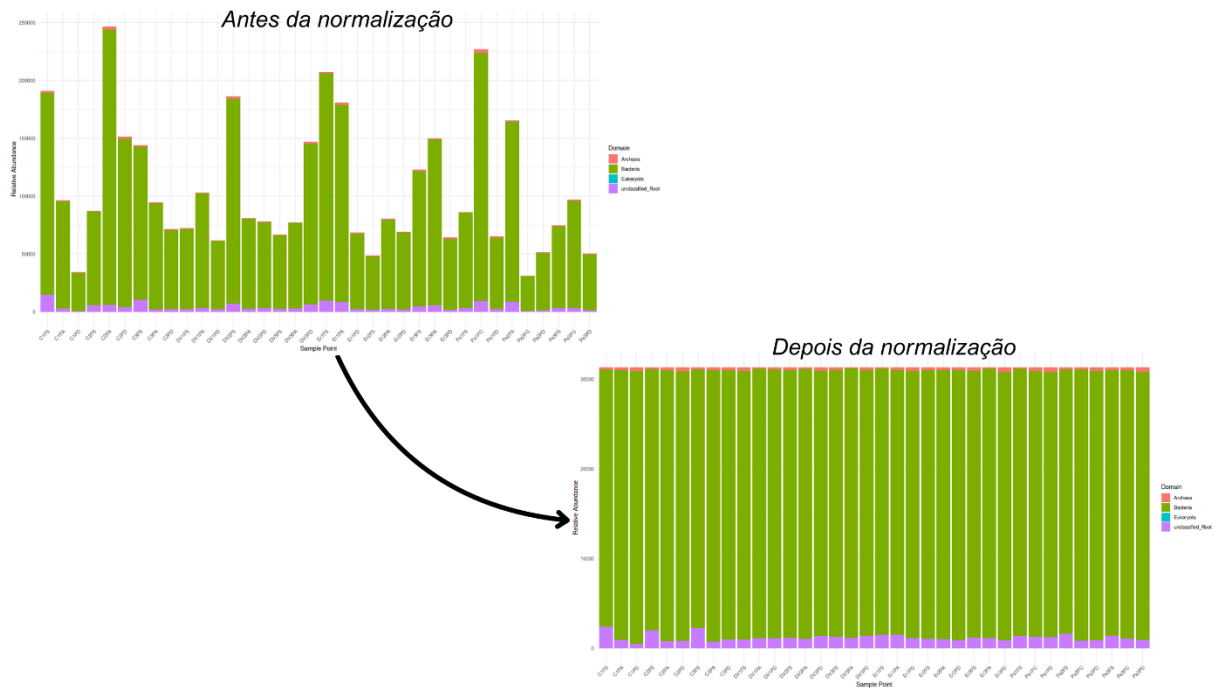
Fonte: O autor.

Como o foco das modelagens preditivas deste estudo recai exclusivamente sobre a comunidade bacteriana, aplicou-se um filtro taxonômico para a remoção de ruídos e redução da dimensionalidade do conjunto de dados. Sendo assim, foram excluídas as ASVs correspondentes aos domínios *Archaea* e *Eukaryota*, bem como sequências não classificadas no nível de domínio. A remoção dessas sequências não classificadas é fundamental, pois elas geralmente representam amplificações inespecíficas ou artefatos gerados durante o PCR (como as quimeras de DNA), cuja permanência inseriria falsos positivos na matriz de dados.

Para mitigar a alta esparsidade típica de dados de sequenciamento e otimizar o desempenho dos algoritmos, adotou-se um critério de filtragem rigoroso, baseado em prevalência e abundância (Cao et al., 2021; Risely et al., 2021; Davies et al., 2022; Siddiqui et al., 2022; Nikodemova et al., 2023; Zhou et al., 2023; Tom et al., 2024). O critério de prevalência exigiu que uma ASV estivesse presente em, no mínimo, duas das três repetições biológicas de um determinado tratamento e localidade. Esse limite garante que a ASV representa uma população residente no solo, descartando detecções transientes ou possíveis contaminações cruzadas de laboratório. Simultaneamente, aplicou-se um critério de abundância absoluta mínima de 50 leituras em pelo menos uma das amostras. Esse limite atua como um *cutoff* computacional para remover erros de sequenciamento (ruído de fundo) e microrganismos ultrarraros, que aumentam drasticamente o número de colunas (variáveis) na matriz sem fornecer poder preditivo aos modelos estatísticos. Após a aplicação de ambos os filtros, obteve-

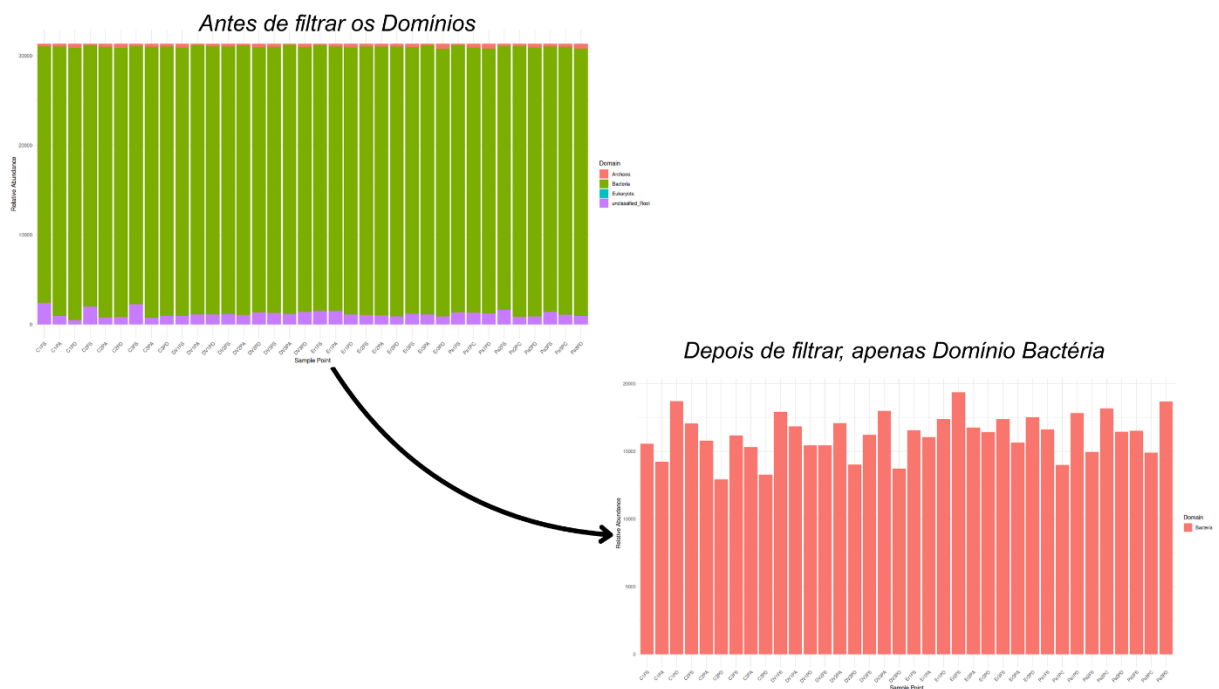
se uma matriz de contagem purificada, restrita a indivíduos consistentes do domínio Bacteria (Figura 8).

Figura 7 - Análise gráfica normalização da abundância de ASVs



Fonte: O autor.

Figura 8 - Análise gráfica filtrando abundância de ASVs para apenas domínio Bacteria



Fonte: O autor.

3.3.4 Ambiente Computacional

O ambiente computacional incluiu o ambiente de desenvolvimento integrado (IDE) RStudio em linguagem R, com bibliotecas para processamento de sequências (DADA2, Biostrings, DECIPHER, SRS), ecologia microbiana e transformação de dados (vegan 2.6-10; dplyr; tidyverse; etc.), e AM (*randomForest* e *xgboost*). Para garantir que a sequência de números aleatórios gerada pelos códigos seja sempre igual, entre servidor e máquina local, foi necessário definir uma semente (função “set.seed(21)”). Em Python foram usadas as bibliotecas “pandas” e “Bio.SeqIO”. A predição funcional por PICRUSt2 (v. 2.6.0) foi executada via conda (Anaconda Distribution). Utilizado software Gephi 0.10.1 para elaboração visual de redes de correlação.

3.4 ALGORITMOS E BIBLIOTECAS ESTATÍSTICAS UTILIZADAS

3.4.1 Análise de Diversidade Alfa

Para caracterizar a diversidade da comunidade bacteriana, a análise de diversidade alfa (riqueza e diversidade dentro das amostras), é fundamental para compreender como diferentes manejos do solo afetam a estrutura das comunidades microbianas. A diversidade alfa foi avaliada através de três métricas calculadas para cada amostra: o índice de riqueza que é o número total de ASVs únicos (função “*specnumber()*” do pacote *vegan*), por Shannon onde mede a diversidade considerando tanto a riqueza quanto a uniformidade das abundâncias (“*diversity()*” com `index = "shannon"`) e por Simpson onde mede a dominância na comunidade (“*diversity()*” com `index = "simpson"`). Foi realizado em sequência, testes de Kruskal-Wallis (não paramétrico) para verificar diferenças significativas nos índices de diversidade entre os tratamentos. Para variáveis com p-valor < 0,05, foi aplicado o teste post-hoc de Dunn com correção de Benjamini-Hochberg, para comparações múltiplas.

Para complementar a análise de diversidade alfa e obter estimativas da riqueza bacteriana, foi realizado uma série de análises utilizando estimadores não paramétricos de riqueza (Chao et al., 2014; Willis, 2019). Utilizando a mesma tabela de contagem de ASVs, a transformação dos dados incluiu conversão para formato de presença/ausência quando necessário, seguindo recomendações de Colwell et al. (2012) para cálculos de estimadores baseados em incidência. Foi utilizado cinco estimadores não paramétricos de riqueza. A riqueza observada (contagem simples de ASVs) serviu como baseline, enquanto os estimadores Chao1 e Chao2, baseados na abundância de espécies raras (Chao et al., 1987), foram calculados para

corrigir a subamostragem. O ACE (*Abundance-based Coverage Estimator*), foi incluído por considerar tanto a cobertura amostral quanto a abundância de espécies raras. Implementado também os estimadores Jackknife de primeira e segunda ordem (Burnham et al., 1978, 1979).

A comparação entre tratamentos foi realizada através de testes de Kruskal-Wallis para cada estimador, seguindo recomendações de McMurdie & Holmes (2014) para análise de dados microbianos com distribuição não-normal. A análise de correlação entre estimadores incluiu linha de identidade e ajuste linear, método validado por Willis & Bunge (2015) para avaliação de desempenho de estimadores. Para visualização, múltiplas abordagens gráficas baseadas em práticas recomendadas por Weiss et al. (2017). Incluí gráficos comparativos entre estimadores, representações individuais para cada método, e tratamento especial para Chao2. Os resultados foram apresentados visualmente por meio de gráficos de boxplot, construídos com o pacote *ggplot2*.

3.4.2 Análise de Diversidade Beta

A Análise de Coordenadas Principais (PCoA) foi utilizada para investigar a diversidade beta, ou seja, para determinar como a composição da comunidade bacteriana varia entre os diferentes tratamentos e locais. Para complementar a análise foi necessário incluir as colunas de diversidade alfa (Shannon, Simpson e Pielou) na tabela de atributos do solo, o que permite posteriormente incorporar estas métricas como vetores ambientais na PCoA. A análise, baseada na matriz de dissimilaridade de Bray-Curtis (matriz de distância calculada a partir da função *vegdist*), permite visualizar os relacionamentos entre as amostras em um espaço de baixa dimensão.

A separação visual observada é confirmada pelo teste de Análise de Variância Multivariada Permutacional (PERMANOVA), onde a PERMANOVA foi realizada pela função “*adonis2()*” do pacote *vegan* com 999 permutações. Os vetores das variáveis ambientais, foram montadas através da função “*rda()*” pacote *rda*, e filtradas a partir da função “*envfit()*” do pacote *vegan* onde é testada 999 permutações e escolhidas apenas os atributos com $p < 0,05$.

3.4.3 PCA e Partição de Variância

Para a análise de Partição de Variância (PV), que visa explicar a variação na estrutura da comunidade bacteriana, é fundamental utilizar um conjunto de variáveis explicativas que seja representativo, mas não redundante. O uso de um número excessivo de variáveis,

especialmente aquelas que são altamente correlacionadas entre si (multicolinearidade), pode inflar a variância explicada e gerar resultados imprecisos.

Para realizar a Análise de Componentes Principais (PCA), utilizou-se a função “*PCA()*” da biblioteca *FactoMineR* (Linguagem R). Esta ordenação foi aplicada exclusivamente aos dados edáficos do solo. para confirmar a premissa de que existe uma relação estatisticamente significativa entre o ambiente e a microbiota, justificando assim o prosseguimento para a modelagem preditiva e Partição de Variação, procedeu-se com o Teste de Mantel. Para a execução deste teste, o conjunto de dados foi dividido na construção de duas matrizes de distância entre as amostras: (i) uma matriz de dissimilaridade biológica, calculada a partir da abundância das ASVs bacterianas utilizando o índice de Bray-Curtis (função “*vegdist()*”, argumento *method* = “bray”, pacote *vegan*); e (ii) uma matriz de distância ambiental, calculada sobre os atributos padronizados do solo utilizando a distância Euclidiana (*method* = “euclidean”). A correlação entre essas duas matrizes foi testada por meio da função “*mantel.partial()*”. O teste, baseado em 9.999 permutações, resultou em uma correlação positiva e altamente significativa ($p < 0,001$). Esse resultado validou a hipótese fundamental de que as variáveis ambientais mensuradas exercem, de fato, forte influência estruturante sobre a composição da comunidade bacteriana observada em cada localidade amostral.

Os atributos selecionados a partir da identificação pela PCA, foi necessário identificar quais atributos são colineares entre si, sendo removido por exemplo o atributo Silte (pois a combinação de Areia e Argila forma Silte), o que formou a combinação dos atributos físicos sendo “Água disponível”, “Porosidade Total”, “Densidade”, “Argila”, “Areia”, “RP”, “PMP”, os atributos químicos “pH (CaCL₂)”, “P”, “V%”, “Mg”, “CTC”, “K” e atributos biológicos “COT”, “CBM”, “Respiração”, “Glomalina total”, “Fosfatase ácida”, “Beta Glucosidase”, “Arilsulfatase”, “Urease”.

A função “*varpart()*” do pacote *vegan* permite particionar a variação total observada na matriz de abundâncias de ASVs (response matrix Y) entre diferentes conjuntos de variáveis explicativas (por exemplo: físico-químico, químico e biológico). Onde a saída da função nos apresenta, partições testáveis (ex.: a, c, etc.), correspondendo à variação exclusivamente explicada por cada conjunto, também partições conjuntas (overlap), que representam a variação explicada por múltiplos conjuntos ao mesmo tempo e a fração residual, que captura a variação não explicada pelos modelos.

3.4.4 Algoritmo Random Forest

O *Random Forest* (RF), presente no pacote “randomForest” para a linguagem R, é um algoritmo de aprendizado por *ensemble* que constrói múltiplas árvores de decisão de forma independente (*bagging*) e combina suas previsões por média (regressão) ou votação majoritária (classificação). Cada árvore é treinada com uma amostra *bootstrap* dos dados e com uma seleção aleatória de preditores em cada nó de decisão. Isso reduz o viés e a variância do modelo, além de permitir a estimativa interna do erro via “*out-of-bag*” (*OOB account*). Robusto a *outliers*, captura relações não lineares e interações sem necessidade de transformação dos dados. Onde sua previsão final é a média ou voto majoritário das árvores (Hermans et al., 2020; Wilhelm et al., 2022; Knight et al., 2024; Lei et al., 2024; Roopashree et al., 2024; Alonso-Sarria et al., 2025).

O modelo implementado foi montado da seguinte forma, onde os preditores são os atributos do solo (biológico, físico e químico), a resposta são os dados bacterianos (abundância ou *leituras*), a quantidade de árvores utilizadas foi de 500, e habilitado o parâmetro *importance* (para calcular a importância das variáveis).

O resultado analisado, disponibilizado pelo RF, é o %IncMSE (*Increment in Mean Square Error*), sendo um dos valores mais precisos do RF, avaliando a redução do poder preditivo do modelo em termos de aumento do erro quadrático médio, quando os valores das amostras (*Out-of-bag* – OOB) são permutados para cada variável preditora, enquanto as outras permanecem fixas, sendo quanto maior o valor de %IncMSE maior a importância da variável (Knight et al., 2024; Lei et al., 2024; Roopashree et al., 2024; Alonso-Sarria et al., 2025).

3.4.5 Algoritmo Extreme Gradient Boosting

O XGBoost (*eXtreme Gradient Boosting*), presente no pacote “xgboost” para a linguagem R, é uma implementação eficiente de *gradient-boosted decision trees* (GBDT). Ao contrário do RF, constrói árvores de forma sequencial, em que cada nova árvore corrige os erros residuais da anterior. Inclui regularização para evitar *overfitting*, controlando a complexidade, parâmetros ajustáveis (*nrounds*, *eta*, *max_depth*, *gamma* etc.) e processamento paralelo. Esse método tende a apresentar desempenho superior em muitos cenários tabulares e amplia a interpretação ao incluir o *ranking* por importância variável. É altamente eficiente computacionalmente (em função de otimizações como o *weighted quantile sketch* e a paralelização), o que o torna adequado para grandes volumes de dados (Wilhelm et al., 2022; Lei et al., 2023; Lei et al., 2024; Roopashree et al., 2024).

O modelo implementado foi montado da seguinte forma: os preditores foram os atributos do solo (biológico, físico e químico), a resposta são os dados bacterianos (abundância ou *leituras*). Isto compõem a matriz *xgb.DMatrix*, diferente do RF que possui apenas dois parâmetros de dados (x e y, sendo x os preditores e y a resposta).

Sendo os parâmetros de configuração, *eta* = 0,05 (Taxa de aprendizado), *max_depth* = 4 (Profundidade), *min_child_weight* = 5; *subsample* = 0,7; *colsample_bytree* = 0,7; *gamma* = 0,1 e *nthread* = 4, após essa configuração temos o modelo, composto por *nrounds* = 1000, *early_stopping_rounds* = 50, *watchlist* = list(train = dtrain) e *verbose* = 0.

O resultado analisado, disponibilizado pelo XGBoost, semelhante ao %IncMSE do RF, temos *ScaledGain*, que vem da fórmula $(importance_matrix\$Gain * (100 / sum(importance_matrix\$Gain)))$, sendo a *importance_matrix* proveniente do cálculo do XGBoost de importância ('weight' + 'gain'). Ao final temos o valor de Gain normalizado, semelhante ao RF, representando aumento médio da função objetivo, o que nos permite comparar com o método estatístico Partição de Variância e com o modelo RF, pois possuem o mesmo propósito analítico (importância relativa das variáveis para o modelo).

3.4.6 Rede de Correlação de Spearman

Para investigar as relações entre a comunidade bacteriana e as propriedades do solo, optou-se por realizar uma análise de rede de correlação utilizando o coeficiente de Spearman. Esta abordagem permitiu identificar associações significativas entre ASVs específicas e variáveis ambientais, agrupadas por tipo de tratamento (Floresta Secundária, Plantio Direto e Pastagem), permitindo explorar associações não lineares entre ASVs e variáveis ambientais, e os potenciais bioindicadores funcionais em diferentes sistemas de manejo.

Foram selecionadas 24 variáveis ambientais, agrupadas em categorias (físicas, químicas, biológicas e índices de saúde do solo), todas padronizadas em escala Z-score para garantir comparabilidade. A matriz de correlação foi construída a partir do coeficiente de Spearman, com aplicação de filtros de significância estatística ($|r| > 0,4$ e $p < 0,05$). Para garantir robustez, um teste de permutação com 1000 iterações foi aplicado, descartando correlações não reprodutíveis. As correlações validadas foram utilizadas para construir uma rede em que os nós representam ASVs ou atributos do solo, e as arestas representam correlações significativas. Cada rede foi construída separadamente por tipo de uso do solo (Floresta Secundária, Plantio Direto e Pastagem), a fim de identificar estruturas ecológicas específicas.

A análise foi implementada em R, utilizando diferentes bibliotecas. As etapas de manipulação e organização dos dados foram conduzidas com auxílio do pacote *tidyverse*, por meio das funções “*filter()*”, “*mutate()*” e “*select()*”. Para a padronização dos atributos ambientais foi utilizada a função “*scale()*”, e a transformação Hellinger (quando aplicada) foi feita com “*decostand()*” do pacote *vegan*. As correlações de Spearman e os p-valores associados foram obtidos com “*rcorr()*” do pacote *Hmisc*, enquanto o pacote *igraph* foi utilizado para construir e manipular os grafos de correlação, por meio das funções “*graph_from_data_frame()*”, “*V()*”, “*E()*” e “*write_graph()*”. Para reorganizar os dados em formato longo e consolidar tabelas de nós e arestas, foi empregada a função “*melt()*” da biblioteca *reshape2*. Além disso, foi desenvolvida uma função personalizada (*criar_rede_tratamento()*) que integra todas essas etapas para gerar, validar e exportar automaticamente as redes para cada sistema de manejo, incluindo o teste de permutação com paralelização via *future.apply*. As redes geradas foram exportadas no formato GraphML, permitindo a visualização no Gephi.

A rede traz diversas informações, nas quais é possível realizar análise de módulos que permite identificar a modularidade da rede, ou seja, como a comunidade bacteriana se organiza em módulos funcionais, fundamental para entender sua resiliência, redundância e coesão. Já a identificação de nós centrais ajuda a revelar nós-chave (*keystone* taxa ou atributos) que podem exercer grande influência na estrutura geral da comunidade. Essas análises combinadas aumentam a capacidade de detectar vulnerabilidades da rede (caso os nós centrais sejam perdidos) e a robustez proporcionada por módulos interconectados (Banerjee et al., 2016; Singavarapu et al., 2023; Byers et al., 2025; Cui et al., 2025; Idbella et al., 2025; Xiang et al., 2025).

3.4.7 Rede Espécies Indicadoras

Com o objetivo de identificar micro-organismos indicativos de cada sistema de manejo do solo, foi realizada uma análise baseada no Índice de Valor Indicador (IVI), que permite detectar ASVs significativamente associadas a um determinado tratamento - seja Floresta Secundária, Plantio Direto e Pastagem. A análise foi implementada em linguagem R, com o auxílio da função “*indval()*” do pacote *labdsv*, que calcula o valor de IVI (representa o cálculo do valor do indicador (fidelidade e abundância relativa) de espécies em grupos ou tipos) e seu p-valor associado por meio de testes de permutação. As entradas para essa função foram uma matriz transposta de ASVs (amostras por colunas e ASVs por linhas) e um vetor com os

tratamentos categorizados. Apenas ASVs com $IVI > 0$ e $p < 0,05$ foram consideradas significativas e mantidas para a construção da rede.

A partir dos resultados, foi construída uma rede bipartida entre ASVs e os tratamentos aos quais estão associadas. Para isso, foi utilizada a biblioteca *igraph*, onde os nós representam ASVs ou tratamentos. As conexões (arestas) entre os nós representam a associação estatística e são ponderadas pelo valor do IVI, sendo mais espessas para ASVs com maior fidelidade ao tratamento. O tamanho dos nós também foi ajustado proporcionalmente ao valor de IVI.

Para explorar o aspecto taxonômico dessas ASVs indicadoras, foi realizada uma agregação por filo com base na classificação taxonômica atribuída anteriormente. As contagens de ASVs indicadoras foram somadas por filo dentro de cada tratamento, e posteriormente comparadas com as ASVs não indicadoras. Esse processo permitiu verificar se determinados filos estavam mais representados como indicadores de algum tipo de manejo. Os resultados foram apresentados visualmente por meio de gráficos de pizza, construídos com o pacote *ggplot2*.

4 RESULTADOS E DISCUSSÕES

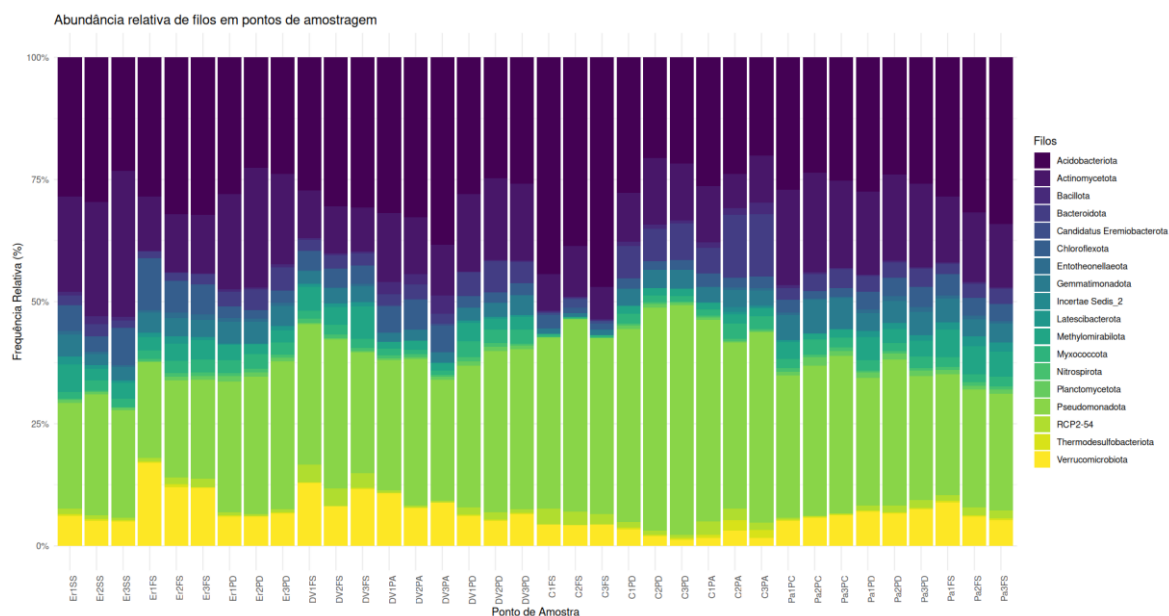
4.1 ANÁLISE DE ABUNDÂNCIA RELATIVA DE FILOS PARA LOCAL E TRATAMENTO

Para compreender como os diferentes usos do solo e as características locais influenciam a estrutura da comunidade bacteriana, foi realizada uma análise do perfil taxonômico em nível de filo (Figura 9), temos na figura cada barra representa um ponto amostral e as cores indicam a abundância relativa (%) dos principais filos identificados. As amostras estão organizadas no eixo X por cidade (C: Castro, DV: Dois Vizinhos, Er: Entre Rios e Pa: Palotina) e por tratamento (FS: Floresta Secundária, PD: Plantio Direto, PA: Pastagem, PC: Preparo Convencional e SS: Sistema Silvopastoril), permitindo a visualização de padrões na composição da comunidade associados a esses fatores. Este tipo de análise é crucial para identificar os principais grupos de microrganismos presentes e observar como sua proporção varia em resposta a diferentes condições ambientais. Ao todo foram classificados taxonomicamente 1023 ASVs de um total de 584.831 leituras (bactérias).

Onde para cada cidade uma quantidade de leituras semelhantes, sendo as amostras da cidade de Entre Rios com 153.081 leituras (26,17%), para a cidade de Dois Vizinhos 144.669 leituras (24,74%), para a cidade de Castro 138.976 leituras (23,76%), e por fim para a cidade de Palotina 148.105 leituras (25,32%).

Em todas as quatro cidades amostradas, a conversão de Floresta Secundária (FS) para uso agrícola (Pastagem ou Plantio) sinaliza aumento de filos copiotróficos em sistemas agrícolas. Essa proporção sugere um ambiente com maior disponibilidade de carbono facilmente assimilável, frequentemente associado à fertilização e à perturbação do solo. Em contrapartida, a redução de grupos oligotróficos, típicos de ambientes estáveis, indica perda de complexidade ecológica, o que pode comprometer processos de longo prazo relacionados à ciclagem de nutrientes e à estabilidade estrutural do solo. (Kim et al., 2021; Zhao et al., 2022; Domeignoz-Horta et al., 2024; Sun et al., 2024).

Figura 9 - Classificação taxonômica das comunidades bacterianas do solo sob diferentes manejos



Fonte: O autor.

Nota-se que no Sistema Silvipastoril (SS), representado pelas amostras Er1SS, Er2SS e Er3SS, verifica-se um perfil taxonômico único que se posiciona com características de transição ecológica entre a floresta e os sistemas agrícolas. Diferente das áreas de manejo agrícola (PA e PD), que mostram uma dominância massiva de *Pseudomonadota*, o Sistema Silvipastoril consegue manter essa população em níveis mais moderados. Ao mesmo tempo, e em contraste com a Floresta Secundária (FS), ele sustenta essa população de *Pseudomonadota* em níveis mais altos, e também retém uma abundância relativa muito maior de filos tipicamente associados à floresta, como *Acidobacteriota* e *Verrucomicrobiota*, em comparação com os outros sistemas agrícolas (Navarrete et al., 2015; Wu et al., 2020; Kim et al., 2021; Liu, W. et al., 2023; Hua et al., 2024).

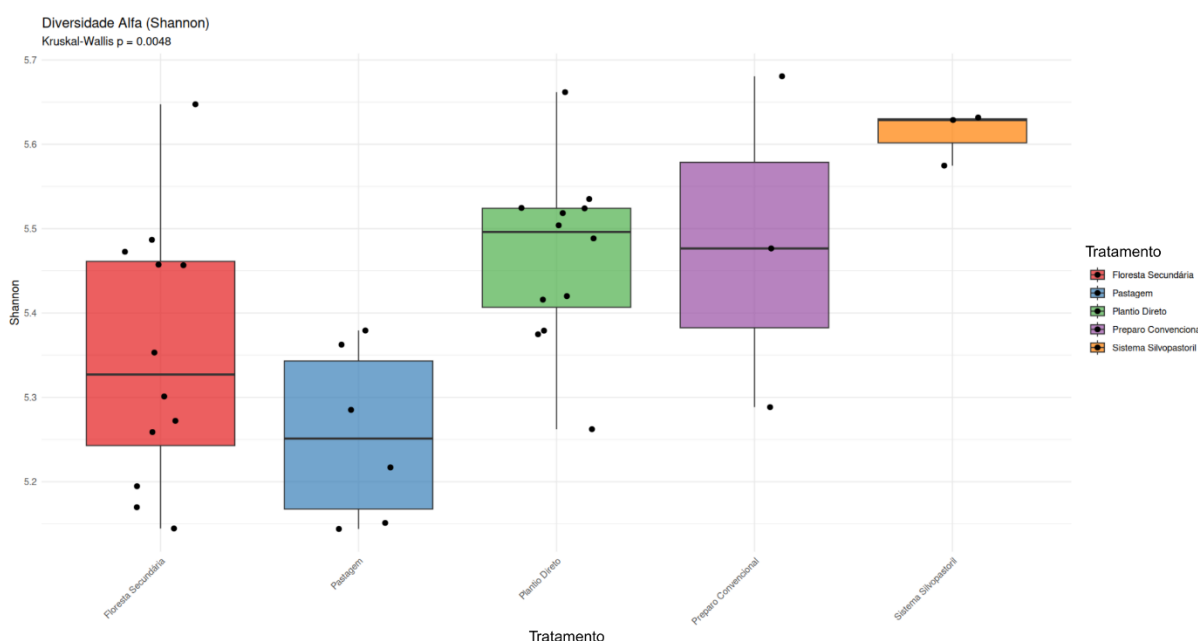
4.2 ANÁLISE DE DIVERSIDADE ALFA DA COMUNIDADE BACTERIANA DOS TRATAMENTOS

A Figura 10, é representada por diagramas de caixa (boxplots) representam a mediana (linha central), os quartis superior e inferior (limites da caixa) e a variação dos dados (hastes). Os pontos individuais representam os valores de diversidade de cada amostra replicada. O valor de p indicado refere-se ao resultado do teste de Kruskal-Wallis, utilizado para comparar as medianas de diversidade entre os tratamentos. Os pontos de coleta relacionados ao plantio direto orgânico foram unidos aos pontos de plantio direto, para facilitar a análise e apresenta os resultados da diversidade alfa, calculada pelo índice de Shannon, que considera tanto a riqueza

(número de espécies ou ASVs) quanto a uniformidade (distribuição de abundância entre elas) em cada amostra, entre os tratamentos: Floresta Secundária, Pastagem, Plantio Direto, Preparo Convencional e Sistema Silvipastoril.

Além da tendência central (mediana), a variabilidade dos dados dentro de cada tratamento também é informativa. O tratamento Pastagem demonstrou a maior amplitude de variação nos valores de diversidade, como indicado pelo tamanho da caixa (intervalo interquartil) e pela dispersão dos pontos, que incluem tanto o valor mais baixo quanto um dos mais altos observados no estudo. Os demais tratamentos apresentaram uma variabilidade interna um pouco menor.

Figura 10 - Diversidade alfa das comunidades bacterianas



Fonte: O autor.

Com o objetivo de avaliar se o manejo do solo impacta a complexidade da comunidade, a diversidade alfa foi calculada utilizando o índice de Shannon (Kruskal-Wallis, $p = 0,0048$). A análise demonstra que os sistemas de manejo apresentam diferentes níveis de diversidade. O Sistema Silvipastoril (SS) apresentou a maior mediana de diversidade alfa, superando todos os demais tratamentos, inclusive a Floresta Secundária (FS). Também é sugerido que sistemas de manejo integrado, que aumentam a complexidade estrutural e a diversidade de insumos de carbono (como a combinação de árvores e pastagem), podem sustentar um maior número de nichos ecológicos, promovendo uma comunidade bacteriana mais rica e uniforme (Carmo et al., 2024; Silva et al., 2024; Funniceili et al., 2025).

Em contraste, a Floresta Secundária apresentou a menor mediana de diversidade, um padrão que pode ser atribuído a fortes filtros ambientais locais (como a extrema acidez do solo em algumas áreas de FS) que limitam a proliferação de táxons sensíveis. Os sistemas convencionais de agricultura (PC, PD, PA) apresentaram níveis de diversidade intermediários. Este resultado indica que, embora a conversão agrícola do solo possa alterar a composição da comunidade, a implementação de sistemas mais complexos, como o silvipastoril, tem o potencial de aumentar ativamente a diversidade microbiana do solo (Carmo et al., 2024; Silva et al., 2024; Funnicelli et al., 2025).

4.3 ESTRUTURA DA COMUNIDADE MICROBIANA ATRAVÉS DA PCOA E NMDS COMPARADA A DBRDA

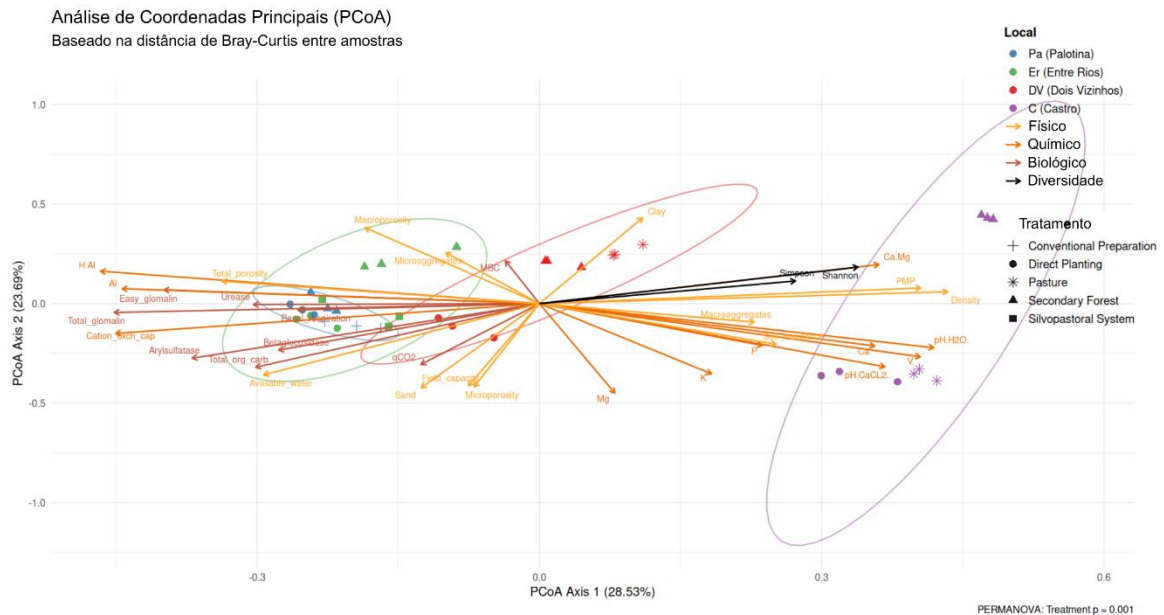
A análise de diversidade beta (PCoA) demonstrou uma forte estruturação da comunidade bacteriana, impulsionada por uma complexa interação entre a biogeografia e o uso do solo (Figura 11), onde a figura representa a Análise de Coordenadas Principais (PCoA) baseada na distância de Bray-Curtis de forma gráfica para a estrutura da comunidade bacteriana (gene 16S rRNA). As cores representam os locais de coleta (C: Castro; DV: Dois Vizinhos; Er: Entre Rios; Pa: Palotina) e as formas indicam os sistemas de uso do solo. As elipses representam o intervalo de confiança de 95% para os grupos de locais. Os vetores indicam as correlações significativas das variáveis edáficas (Físicas, Químicas, Biológicas) e índices de diversidade com os eixos de ordenação. A significância do efeito do tratamento foi confirmada por PERMANOVA ($p = 0,001$). A variação explicada por cada eixo é indicada entre parênteses. O Eixo 1 (28,53%) demonstra um padrão biogeográfico distinto, isolando primariamente as amostras de Floresta Secundária (FS) de Castro (C-FS) das demais. Ao verificar os dados edáficos nota-se que essas amostras C-FS representam um ambiente único, caracterizado por extrema acidez ($\text{pH} \sim 4,15$), o que as diferencia não apenas das outras localidades, mas também das áreas agrícolas da própria Castro, que apresentavam pH mais elevado. Os vetores ambientais (como $\text{pH.H}_2\text{O}$, Ca, P, Density e as métricas de diversidade Shannon/Simpson) mostraram-se correlacionados com este eixo, refletindo a assinatura química e física dos solos agrícolas de Castro, que se distinguem dos solos florestais ácidos.

O Eixo 2 (23,69%) representou o principal eixo de impacto do uso do solo para as demais localidades (Pa, Er, DV), separando significativamente (PERMANOVA, $p = 0,001$) as comunidades de Floresta Secundária das comunidades de uso agrícola. As comunidades florestais associaram-se a indicadores de saúde do solo, como maior *Macroporosity* e MBC

(Biomassa Microbiana), enquanto os solos agrícolas agruparam-se em função de outros atributos. Assim temos que, a PCoA demonstra que, embora o manejo agrícola induza uma mudança consistente (Eixo 2), um filtro ambiental biogeográfico extremo, como a combinação de solo e vegetação nativa em Castro, pode criar uma assinatura microbiana única (Eixo 1) que se sobrepõe ao efeito geral do tratamento. Embora a conversão agrícola cause uma mudança taxonômica consistente (o *turnover* de *Acidobacteriota* para *Pseudomonadota*, como visto na análise de filós), a composição basal da comunidade é, em primeiro lugar, definida por fortes filtros biogeográficos e ambientais específicos de cada localidade, como o local da cidade de Castro.

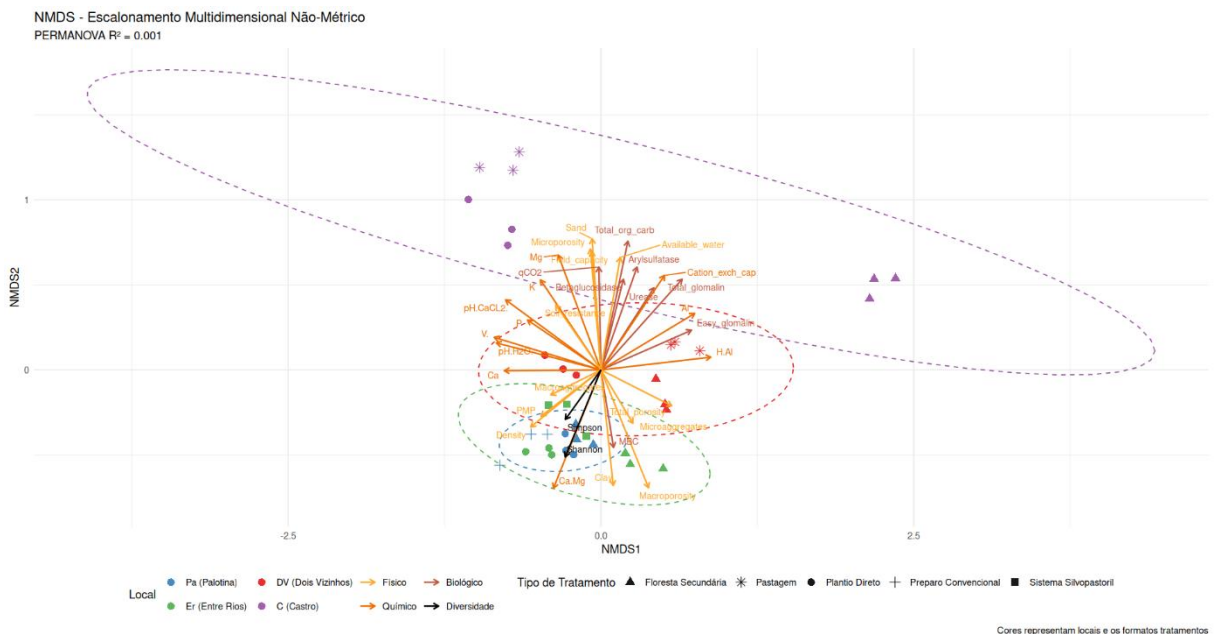
Para validar a estrutura da comunidade observada na PCoA e confirmar os padrões ecológicos, um escalonamento multidimensional não-métrico (NMDS) foi realizado (Figura 12), onde a figura representa a Ordenação por Escalonamento Multidimensional Não-Métrico (NMDS) da comunidade bacteriana do solo graficamente, baseada na dissimilaridade de Bray-Curtis. As cores dos símbolos representam os locais de coleta e as formas indicam os sistemas de uso da terra. As elipses (nível de confiança de 95%) delimitam os centróides dos locais. Os vetores coloridos representam as variáveis ambientais (Físicas, Químicas, Biológicas) e índices de diversidade que apresentaram correlação significativa ($p < 0,05$) com a ordenação, onde o comprimento da seta é proporcional à força da correlação. A análise de PERMANOVA indicou diferenças significativas na composição da comunidade entre os grupos ($p = 0,001$). Além disso, uma análise de Procrustes foi usada para quantificar a concordância entre a ordenação não-restringida (NMDS, que representa a variação total da comunidade) e uma ordenação restringida (dbRDA/CAP, que representa apenas a variação explicada pelos atributos do solo). A solução de 2 dimensões foi estatisticamente robusta, alcançando um valor de stress excelente de 0,065.

Figura 11 - Análise de Coordenadas Principais (PCoA)



Fonte: O autor.

Figura 12 - Ordenação Não-Métrica NMDS



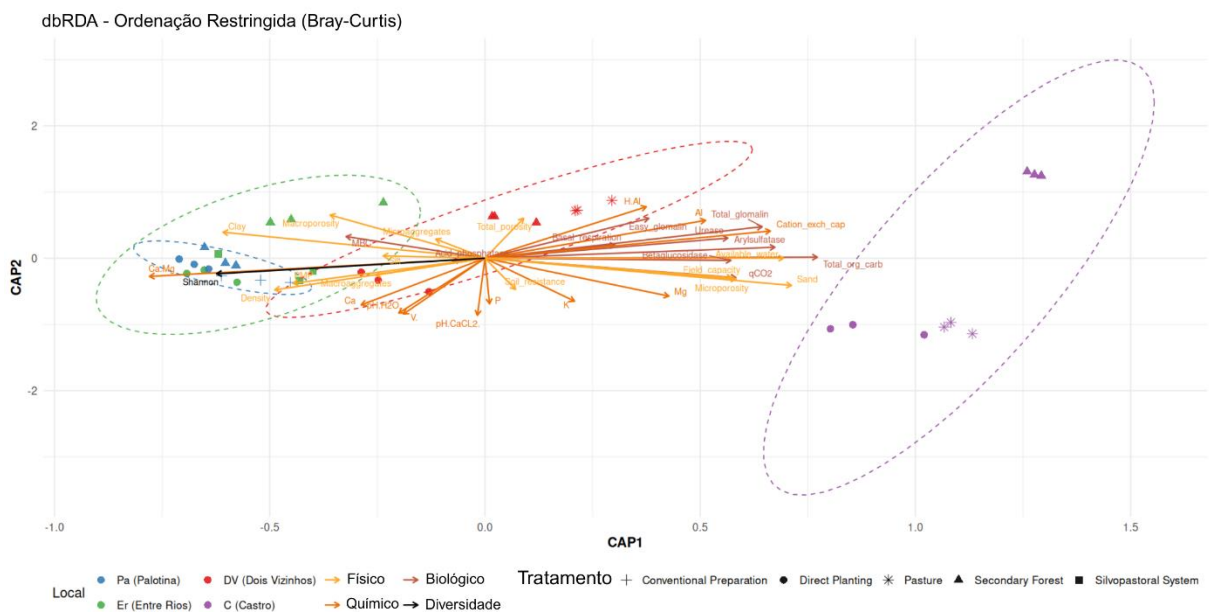
Fonte: O autor.

A convergência de uma análise métrica (PCoA) e uma análise não-métrica (NMDS) no mesmo resultado, em que o NMDS corroborou os resultados da PCoA, demonstrando a mesma hierarquia de filtros, sendo uma separação primária ao longo do Eixo 1 impulsionada por filtros biogeográficos (isolando Castro) e uma separação secundária ao longo do Eixo 2 impulsionada pelo uso do solo (FS vs. Agricultura).

Para quantificar o poder explanatório total dos fatores edáficos medidos, foi realizada uma Análise de Procrustes (*protest*), comparando a ordenação não-restringida (NMDS) com uma ordenação restringida (dbRDA/CAP), Figura 13, onde a figura representa graficamente a Análise de Redundância baseada em distância (dbRDA) visualizando a relação entre a estrutura da comunidade bacteriana e as variáveis ambientais predictoras. A ordenação foi construída a partir da matriz de dissimilaridade de Bray-Curtis. Os eixos canônicos (CAP1 e CAP2) representam a variação linearmente restringida pelos parâmetros edáficos. As cores identificam os locais de amostragem e os formatos indicam os regimes de manejo. Os vetores (setas) mostram a direção e a magnitude da correlação das variáveis físicas, químicas e biológicas com a estrutura da comunidade. As elipses indicam o intervalo de confiança de 95% para os centróides dos locais. A significância global do modelo foi verificada por teste de permutação (ANOVA, $p < 0,05$).

Os resultados indicaram concordância entre os dois modelos ($r = 0,9585$, $p = 0,001$), demonstrando que os atributos físicos, químicos e biológicos do solo medidos são, de fato, os principais fatores determinísticos que governam a assembleia da comunidade bacteriana, com pouca variação estrutural deixada inexplicada.

Figura 13 - Ordenação Restringida dbRDA



Fonte: O autor.

A análise também revelou um forte efeito local, a elipse roxa correspondendo a Castro (altitude ~1025 m) situa-se isolada à direita do gráfico, sugerindo uma comunidade bacteriana distinta, moldada por suas particularidades altitudinais, edáficas e climáticas. Em contraste, as elipses de Palotina, Dois Vizinhos e Entre Rios se sobrepõem à esquerda, indicando maior similaridade entre essas áreas agrícolas, reforçando a influência conjunta do manejo e das condições ambientais locais. Os vetores com direção oposta à esquerda ($H+Al$ $R^2 = 0,96$; carbono orgânico total, $R^2 = 0,79$; glomalina total e facilmente extraível, $R^2 = 0,82$ e $0,63$; água disponível, $R^2 = 0,84$; atividades enzimáticas como arilsulfatase e β -glicosidase, $R^2 = 0,84$ e $0,53$) estão ligados aos solos agrícolas de Palotina, Dois Vizinhos e Entre Rios, indicando solos com maior acidez potencial, atividade microbiana e riqueza em carbono e glomalina (Domeignoz-Horta et al., 2024; Tapaça et al., 2024; Yoon et al., 2024).

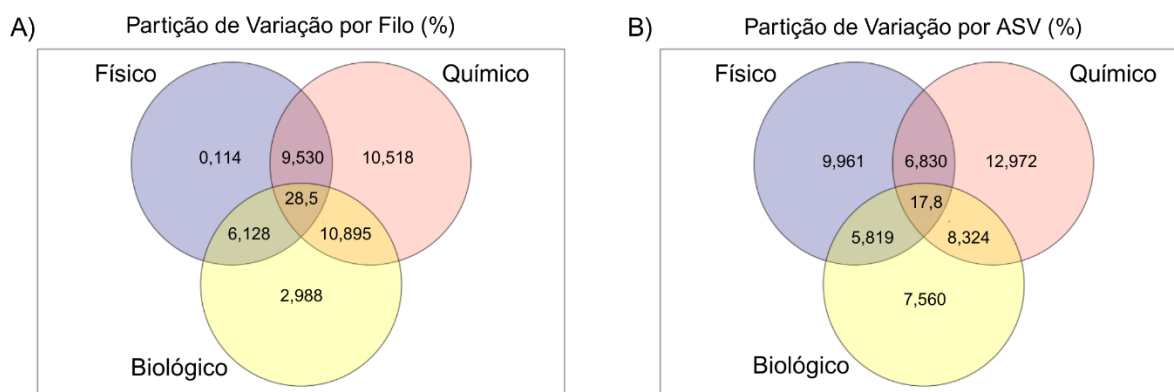
O fator mais dominante é o uso do solo, que promove uma mudança significativa na composição microbiana ao converter a FS para a agricultura. Esta mudança está intrinsecamente ligada a um conjunto de alterações nos atributos do solo, enquanto a FS preserva características de maior acidez, os sistemas agrícolas, embora apresentem uma tendência a maior diversidade alfa e atividade enzimática, são caracterizados por menor acidez e uma comunidade microbiana funcionalmente distinta (Zhang et al., 2022; Mo et al., 2022; Wang; Li, 2023; Reis et al., 2025). Em um segundo nível, a análise evidencia um forte sinal biogeográfico, a cidade de Castro abriga uma comunidade bacteriana distinta, associada a solos de maior pH em local de floresta secundária, enquanto as cidades de Palotina, Dois Vizinhos e Entre Rios compartilham um perfil microbiano mais similar entre si em seus ambientes agrícolas, associado a maior acidez e níveis de carbono orgânico (Bay et al., 2020; Bai et al., 2022; Goss-Souza et al., 2022). Assim, a análise revela que tanto a diversidade alfa quanto a composição da comunidade (abundância relativa) são profundamente moduladas pelo manejo e contexto edafoclimático (Zheng et al., 2019; Liu et al., 2022; Aqeel et al., 2024; Qi et al., 2025), permitindo compreender os impactos do uso da terra sobre a microbiota do solo de forma integrada.

4.4 ANÁLISE PARTIÇÃO DE VARIAÇÃO

Com o objetivo de selecionar os atributos edáficos que influenciam a composição da comunidade bacteriana (ASVs), é aplicado uma Análise de Componentes Principais (PCA) sobre estes dados. Onde a abordagem escolhida foi de realizar individualmente uma PCA para cada tipo de conjunto de atributos, tal que na Figura 14, detalha-se a PCA para os três conjuntos de atributos do solo: (A) Físicos, (B) Químicos e (C) Biológicos. Os gráficos de correlação

ecológico em que as características físicas, químicas e biológicas atuam de forma indissociável. Entre as frações puras, os atributos Químicos emergiram como o filtro ambiental independente mais influente (fração [b] = 10,5%), enquanto as contribuições isoladas dos componentes Biológico ([c] = 3,0%) e Físico ([a] = 0,1%) foram marginais. Ao aumentar a resolução para o nível de ASV (Figura B), observou-se uma mudança notável na arquitetura da variância explicada. A fração compartilhada global [g] diminuiu para 17,8%, permitindo que os efeitos dos filtros ambientais específicos se tornassem mais distintos. O componente Químico manteve sua primazia como o principal preditor puro ([b] = 13,0%), sugerindo que o pH e a disponibilidade de nutrientes (ex: P, K, Mg) exercem uma pressão seletiva direta e insubstituível na seleção de espécies.

Figura 15 - Análise de Partição de Variação através de Diagramas de Venn



Fonte: O autor.

Paralelamente, a influência pura dos atributos Físicos aumentou drasticamente de 0,1% (Filos) para 10% (ASVs), indicando que a estrutura física do solo (porosidade, densidade) atua como um filtro de habitat crítico para a seleção de variantes específicas, embora não afete a composição em níveis taxonômicos superiores. O componente Biológico puro explicou 7,6% da variação. É importante ressaltar que, embora a fração pura biológica pareça menor que a química na VarPart (linear), a alta proporção de variância compartilhada ([g]) sugere que a biologia do solo está intrinsecamente ligada à modulação das propriedades físico-químicas, atuando como um engenheiro do ecossistema cuja influência permeia todo o modelo.

Os resultados indicam uma hierarquia clara de filtragem ambiental. Enquanto os atributos químicos (ex.: pH, disponibilidade de nutrientes) atuam como restrições ecológicas universais, exercendo a maior influência seletiva direta tanto em níveis taxonômicos superiores (Filos) quanto em alta resolução (ASVs), a estrutura física do solo (ex.: macroporosidade, densidade) emerge como um filtro crítico especificamente na escala de ASVs. Sugerindo que,

após a definição química do nicho, a arquitetura física do solo modula a coexistência microbiana por meio da provisão de micro-habitats (Labouyrie et al., 2023; Meszarosova et al., 2024; Sun et al., 2024).

Contudo, a magnitude da variação compartilhada (intersecção) em ambos os modelos reforça a visão do solo como um sistema integrado, em que nenhuma propriedade atua de forma totalmente isolada. A influência do componente biológico, embora menor como fração pura, manifesta-se vigorosamente por meio de sua interação com as propriedades físico-químicas, sugerindo que a biota residente participa ativamente da manutenção da estrutura e da fertilidade do solo (Wang et al., 2019; Labouyrie et al., 2023; Sun et al., 2024; Yan et al., 2025).

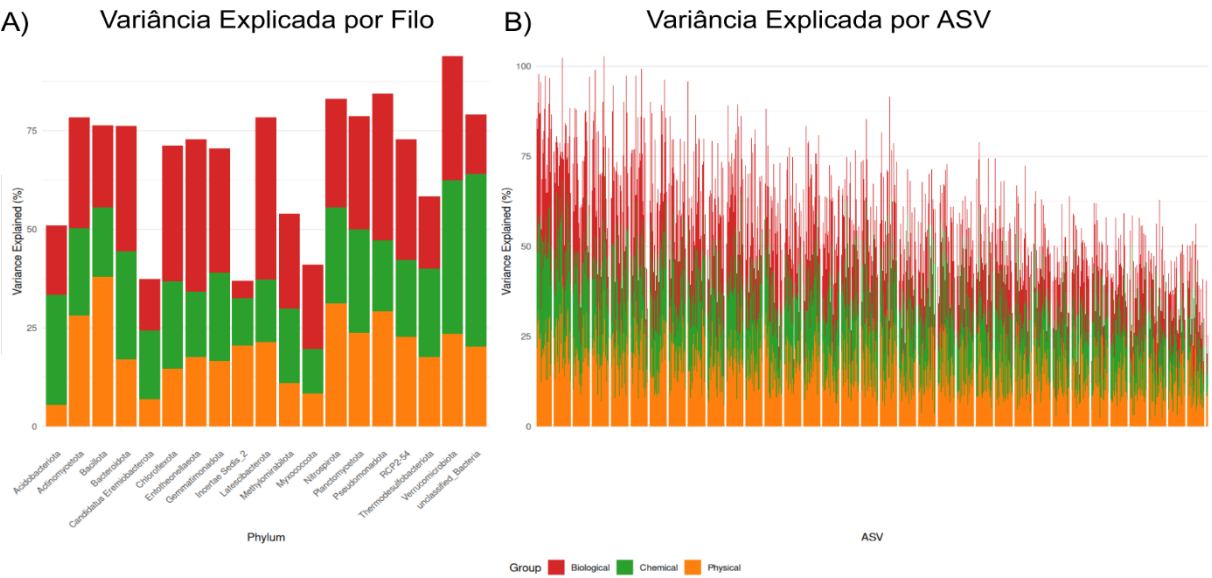
Portanto, a PV demonstra que a predição da estrutura da comunidade bacteriana depende primariamente de filtros edáficos abióticos (químicos e físicos). No entanto, a alta conectividade entre as variáveis (fração compartilhada) evidencia que a quantificação isolada desses fatores é insuficiente. É a interação dinâmica entre os limites abióticos e o capital biológico residente (enzimas, biomassa) que define o nicho realizado, atuando como um mecanismo co-determinante na assembleia da comunidade em nível de ASV.

4.5 ESTRUTURAÇÃO DA COMUNIDADE BACTERIANA ATRAVÉS DE ALGORITMOS DE MACHINE LEARNING

Para complementar e validar os resultados da PV, foi empregada uma abordagem analítica distinta, baseada em AM. Enquanto a PV é fundamentada em ordenação linear (RDA) e sensível à multicolinearidade (exigindo uma pré-seleção de variáveis via PCA), os algoritmos de AM baseados em árvores (*Random Forest*, *XGBoost* - *eXtreme Gradient Boosting*, *Gradient Boosting Machine*) são intrinsecamente projetados para capturar interações complexas e não-lineares, oferecendo uma estrutura independente para testar as mesmas hipóteses. Diferente da PV, que avalia a comunidade como um todo, esta abordagem cria um modelo preditivo para cada filo (Figura 16A, Figura 17A e Figura 18A) e cada ASV (Figura 16B, Figura 17B e Figura 18B) individualmente. Os atributos utilizados para esta análise são os mesmos obtidos pela seleção da PCA do capítulo anterior. Na Figura 16, temos os gráficos de importância de variáveis derivados de modelos de *Random Forest*, mostrando a porcentagem de variância explicada para cada táxon bacteriano. O painel da esquerda (A) exibe a variância explicada para cada Filo, enquanto o painel da direita (B) mostra a mesma análise para cada ASV individual. A altura total de cada barra representa a capacidade preditiva do modelo, métrica %IncMSE - *Increment in Mean Square Error* para *Random Forest* (RF), métrica *Gain* para *XGBoost*,

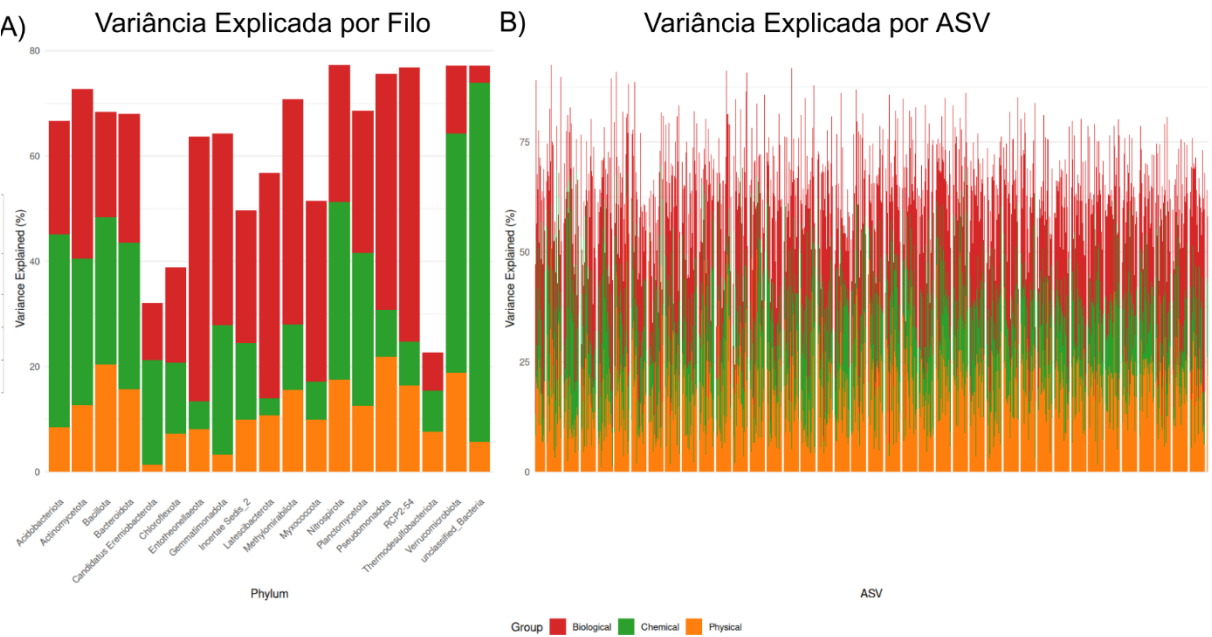
métrica *rel.inf* - *Relative Influence* para *Gradient Boosting* (GB), para aquele táxon específico. As cores empilhadas indicam a importância relativa dos três grupos de preditores ambientais (Biológico, Químico e Físico) na explicação da abundância de cada táxon.

Figura 16 - Gráficos de importância de variáveis derivados de modelos de Random Forest



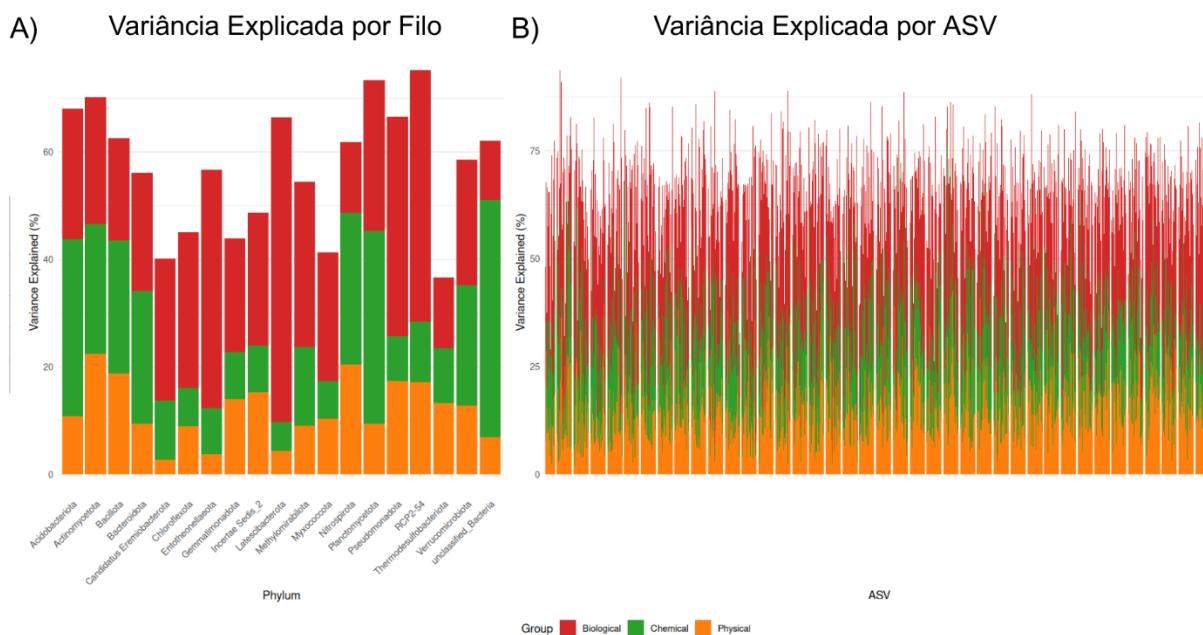
Fonte: O autor.

Figura 17 - Gráficos de importância de variáveis derivados de modelos de XGBoost



Fonte: O autor.

Figura 18 - Gráficos de importância de variáveis derivados de modelos de GBM



Fonte: O autor.

A altura total das barras indica que a abundância de alguns filos é bem explicada pelo modelo, enquanto a de outros é menos previsível. Filos como *Verrucomicrobiota*, *Pseudomonadota*, *Nitrospirota* e *Actinomycetota* apresentam altos valores de variância explicada ($\%IncMSE > 75\%$), indicando que suas populações são fortemente estruturadas pelos fatores ambientais medidos. A observação mais importante é a composição das barras. Para a grande maioria dos filos, o segmento Biológico é o maior, indicando que os atributos biológicos do solo são o grupo de preditores mais importante (Tabela 2). Apesar da enorme variabilidade entre os milhares de ASVs, o atributo Biológico domina a composição da vasta maioria das barras. Demonstrando que a importância dos fatores biológicos não é apenas um artefato de agregação em nível de filo, mas uma regra que se aplica à maioria dos táxons individuais que compõem a comunidade (Wang et al., 2019; Devkota et al., 2024; Liao et al., 2024).

A previsibilidade (altura das barras) para ASVs individuais é, em média, menor e muito mais variável do que para filos. É inerentemente mais difícil prever a abundância de uma única espécie do que a de um grande grupo taxonômico. A inferência de que o componente biológico é a principal categoria que influencia a estrutura da comunidade, é demonstrada pelos resultados dos algoritmos de AM, mesmo que utilizem abordagens estatísticas fundamentalmente diferentes. Enquanto a PV mostrou a importância geral dos fatores para a comunidade como um todo, os algoritmos oferecem um nível de detalhe maior. O que permite afirmar que não apenas o componente biológico é importante de forma geral, mas ele é o mais importante para

quase todos os filos específicos e para a maioria das ASVs individuais. Ele também nos permite identificar quais táxons podem ser exceções a essa regra, sendo mais influenciados pela química ou física do solo. A combinação das análises de PV e AM, observa-se que as condições biológicas do solo, refletidas pela atividade enzimática, presença de glomalina e outros indicadores do legado microbiano, são os preditores mais consistentes e importantes da estrutura da comunidade bacteriana neste sistema, desde o nível de Filo até o de ASVs.

Em nível de Filo, os modelos alcançaram um alto grau de acurácia, com R^2 de validação cruzada (R^2 CV médio) variando de 0,642 - GB a 0,718 - RF (Tabela 3). Verifica-se que na granularidade extrema do nível de ASV, onde os processos estocásticos são altos, os modelos mantiveram uma alta performance preditiva (R^2 CV médio de 0,536 a 0,570), confirmando a força do sinal ambiental determinístico.

Apesar das arquiteturas de modelo fundamentalmente diferentes (*bagging* no RF e *boosting* no XGB/GB) e da abordagem estatística distinta da PV, todos os métodos identificaram, o mesmo padrão, onde o conjunto de atributos Biológicos (vermelho) é o principal contribuinte para a predição da abundância taxonômica, seguido pelos Químicos (verde) e Físicos (laranja). Esta paridade de resultados por métodos independentes confere um alto grau de confiança ao estudo, demonstrando que esta hierarquia é um sinal ecológico genuíno e não um artefato de uma única abordagem analítica.

Observa-se que os principais filos, envolvidos na mudança do uso do solo de floresta secundária e agricultura, são também os mais previsíveis. *Pseudomonadota* e *Acidobacteriota*, juntamente com *Myxococcota*, apresentam valores de R^2 consistentemente altos (próximos de 0,9 no R^2 médio de treino e validados pelo R^2 CV de ~0,7), isso sugeriria que estes filos são bons candidatos a indicadores de mudança no uso do solo. Tal resultado indica também que a abundância desses filos-chave não é aleatória; ela é fortemente governada pelos fatores edáficos medidos. Em contrapartida, filos como *Bacteroidota* e *Desulfobacterota* mostraram baixa previsibilidade (barras de R^2 baixas), sugerindo que suas populações são controladas por processos estocásticos ou por variáveis bióticas não mensuradas.

A variação na abundância de *Pseudomonadota* é explicada quase exclusivamente pelos atributos Biológicos (barra predominantemente vermelha). O que pode indicar um forte acoplamento e uma interação bidirecional, essas bactérias não apenas respondem à disponibilidade de substratos orgânicos, mas também são, muito possivelmente, as principais produtoras das enzimas extracelulares e as maiores constituintes do Carbono da Biomassa Microbiana (MBC) mensurados. Sendo assim, o modelo de aprendizado de máquina capturou com essa covariância biológica. Em contraste, filos como *Acidobacteriota* e *Verrucomicrobiota*

exibem uma assinatura de controle mista, onde o componente Químico (verde) demonstra um peso explicativo quase tão relevante quanto o Biológico. Esse padrão visto indica que esses grupos oligotróficos são reconhecidos por sua extrema sensibilidade ao pH do solo. Nesse caso, a química do solo atua como um filtro ambiental primário e unidirecional, co-limitando o estabelecimento dessas populações independentemente das flutuações na biomassa total do ecossistema.

Esta abordagem demonstrou alta previsibilidade ($R^2 > 75\%$) dos principais filos envolvidos no turnover ecológico (*Pseudomonadota* e *Acidobacteriota*), confirmando que sua alternância é determinística e está ligada a fatores edáficos. E a existência de assinaturas de controle distintas. A abundância de *Pseudomonadota* foi predominantemente associada a fatores biológicos, enquanto *Acidobacteriota* e *Verrucomicrobiota* responderam a um controle misto, químico-biológico, no qual o pH provavelmente atua como um co-limitante. Este método de uso de algoritmos de AM, portanto, auxilia a estatística tradicional, mas também a enriquece, detalhando quais táxons são previsíveis e como são controlados.

Tabela 1 - Comparação valores preditos por VarPart, RF, XGBoost e GB para atributos biológicos, químicos e físicos e residual

MÉTODO	NÍVEL TAXONÔMICO	BIOLÓGICO (%)	QUÍMICO (%)	FÍSICO (%)	RESIDUAL (%)
VarPart	Filo	2,98	10,51	0,11	31,31
RF (%IncMSE)	Filo	27,92	21,04	18,53	32,49
XGBoost (Gain)	Filo	29,27	23,42	12,77	34,54
GB (rel.inf)	Filo	29,46	20,07	14,58	35,87
Varpart	ASV	7,56	12,97	9,96	30,74
RF (%IncMSE)	ASV	25,17	22,04	19,82	32,96
XGBoost (Gain)	ASV	27,49	17,71	16,76	38,03
GB (rel.inf)	ASV	30,12	18,44	14,27	37,16

Fonte: O autor.

Tabela 2 - Valores explicativos e de validação dos algoritmos de aprendizado de máquina

ALGORITMO	NÍVEL TAXONÔMICO	R ² MÉDIO	RMSE MÉDIO	R ² CV MÉDIO
RF (%IncMSE)	Filo	0,940	0,534	0,718
XGBoost (<i>Gain</i>)	Filo	0,953	0,123	0,690
GB (rel.inf)	Filo	0,973	0,149	0,642
RF (%IncMSE)	ASV	0,867	8,304	0,570
XGBoost (<i>Gain</i>)	ASV	0,878	7,740	0,548
GB (rel.inf)	ASV	0,907	6,481	0,536

Fonte: O autor.

4.6 ANÁLISE REDE DE CORRELAÇÃO SPEARMAN PARA USOS DO SOLO

Para compreender as interações específicas entre as variáveis ambientais e os táxons microbianos individuais no ecossistema de referência, foi construída uma rede de correlação para o tratamento Plantio Direto (Figura 19), Floresta Secundária (Figura 20) e Pastagem (Figura 21). Esta análise permite visualizar quais fatores do solo atuam como os principais nós centrais ou moduladores da comunidade bacteriana neste ambiente preservado. Onde nas três figuras (Figura 19, Figura 20 e Figura 21), os nós maiores representam os atributos do solo, coloridos de acordo com sua categoria funcional: Azul (Físicos), Verde (Químicos) e Vermelho (Biológicos). Os nós menores representam as ASVs. As arestas indicam correlações estatisticamente significativas ($|\rho| > 0,4$; $p < 0,05$). A estrutura da rede e os agrupamentos visuais destacam os nichos ecológicos que organizam a comunidade bacteriana.

A rede da Floresta Secundária (Figura 19) exibe uma topologia específica (número de nós, arestas, densidade) que reflete a ecologia deste ambiente, a rede é representada por 522 nós, sendo 21 nós representando os atributos físico-químicos e biológicos do solo, tendo 501 ASVs ligadas a esta estrutura, e 882 arestas (conexões entre ASVs e atributos), tal que a abundância total desta rede é de 132.937 leituras (22,73%). A rede de correlação do tratamento Pastagem (Figura 20) revela uma arquitetura de interações complexa e distinta, que caracteriza um ecossistema agrícola de alta produtividade, a rede é representada por 481 nós, sendo 21 nós

representando os atributos físico-químicos e biológicos do solo, tendo 460 ASVs ligadas a esta estrutura, e 1523 arestas, tal que a abundância total desta rede é de 103.001 leituras (17,61%). A rede de correlação do tratamento Plantio (Figura 21) exibe uma topologia notavelmente organizada e esteticamente distinta, a rede é representada por 448 nós, sendo 21 nós representando os atributos físico-químicos e biológicos do solo, tendo 427 ASVs ligadas a esta estrutura, e 821 arestas (conexões entre ASVs e atributos), tal que a abundância total desta rede é de 159.757 leituras (27,31%). Esta organização visual sugere um ecossistema governado por filtros ambientais fortes e bem definidos, onde diferentes guildas microbianas se associam a eixos funcionais específicos.

A análise de correlação para o tratamento PD (Figura 19) revela uma rede altamente modularizada, indicando que a assembleia da comunidade é governada por três conjuntos de filtros de nicho distintos. O módulo mais denso e central é o Biológico, onde MBC e glomalina total atuaram como nós centrais primários, indicando uma forte auto-organização da comunidade em torno do capital biológico residente e da atividade enzimática. Além disso, demonstra-se também que a principal força que estrutura a comunidade no sistema de PD é um ciclo de retroalimentação biológica positiva (Costa et al., 2024; Ma, L. et al., 2024). Um grande conjunto de ASVs está positivamente correlacionado com a biomassa microbiana e fúngica (Glomalina) já existente, sugerindo que a formação biológica do solo (a biomassa residente) gera nichos e fornece substratos que selecionam uma guilda específica de ASVs. Por sua vez, muitas dessas ASVs provavelmente contribuem para a atividade enzimática (como Betaglucosidase), quebrando matéria orgânica e reabastecendo o MBC, criando um ciclo que se auto reforça (Domeignoz-Horta et al., 2024; Ma, L. et al., 2024).

Em segundo lugar, o módulo físico atuou como um filtro de habitat, segregando guildas de ASVs que respondem positivamente à porosidade total versus aquelas associadas à densidade e argila. O sistema de Plantio Direto não é homogêneo. Ele apresenta uma variação interna na física do solo (algumas áreas mais compactadas, outras menos). A rede indica que essa variação física atua como um filtro de seleção secundário, separando a comunidade em especialistas de habitat. Enquanto o módulo Químico distinto, centrado no pH e nos nutrientes (P, K), representando a guilda de especialistas que respondem diretamente às práticas de manejo de fertilidade, como a adubação química (Huang et al., 2019). As ASVs conectadas a este módulo são aquelas que respondem positivamente às práticas de manejo de correção e fertilização do solo, tais como calagem e adição de P e K (Huang et al., 2019). Elas representam a guilda copiotrófica que é diretamente selecionada pela disponibilidade de nutrientes e pela redução da acidez, um pilar fundamental do sistema de plantio, formando uma rede estável

(Cornell et al., 2023; Song et al., 2023; Byers et al., 2024). Esse padrão está de acordo com estudos que mostram que uso agrícola intensivo aumenta a complexidade e conectividade de redes microbianas, mas tende a reduzir a redundância e modularidade (como descrito para solos cultivados intensivamente) (Karimi et al., 2019; Chen, W et al., 2022).

Figura 19 - Rede de correlação de Spearman para o tratamento Plantio Direto



Fonte: O autor.

Figura 20 - Rede de correlação de Spearman para o tratamento Floresta Secundária



Fonte: O autor.

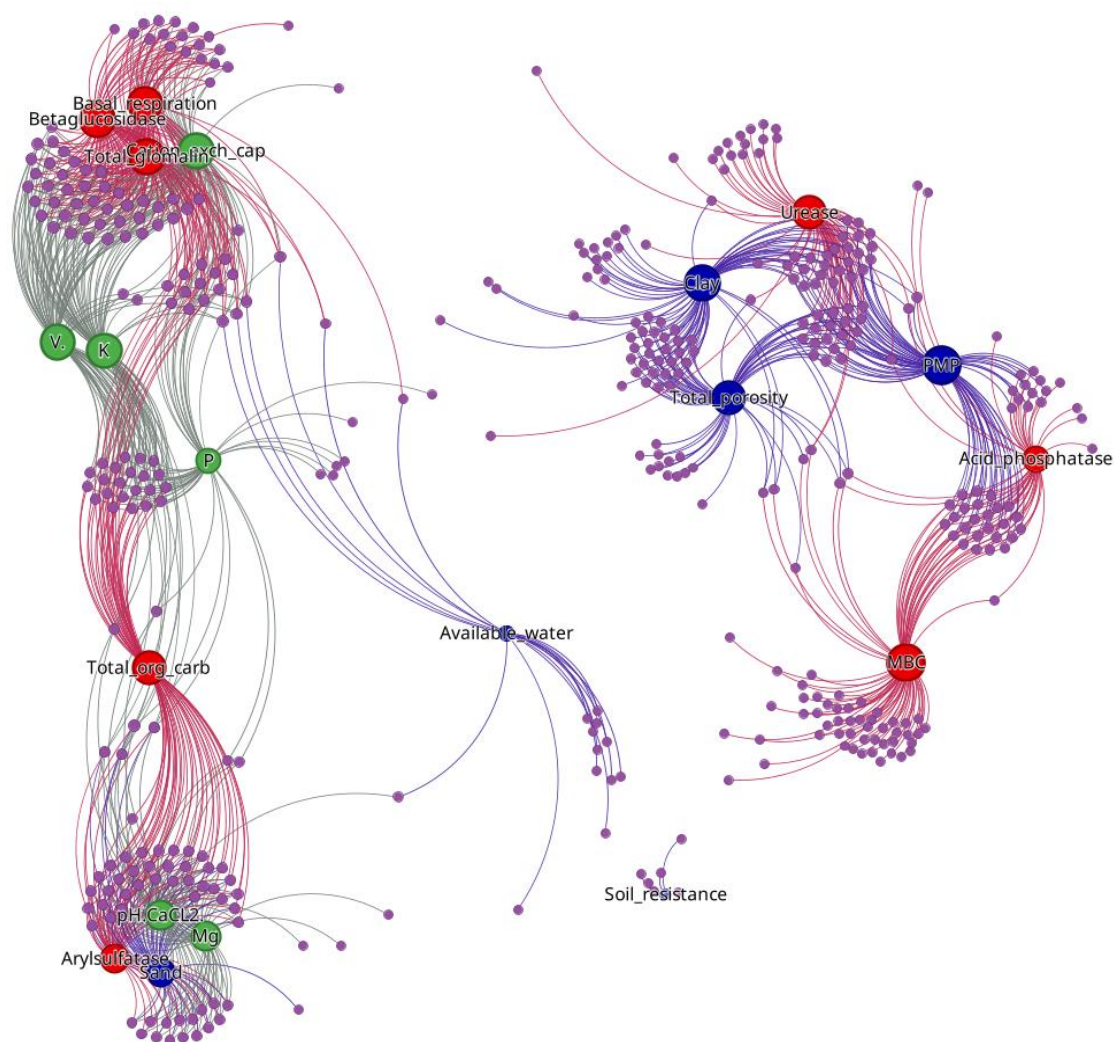
Em nítido contraste com a estrutura modularizada do PD, a rede de correlação da Floresta Secundária (Figura 20) demonstra um forte acoplamento entre os atributos biológicos e químicos. O nó central de atividade biológica (Betaglucosidase, glomalina total) e os nós centrais químicos (pH.CaCL₂, Mg, K) não formaram módulos distintos, mas sim um super módulo bioquímico integrado (Guseva et al., 2022; Yang et al., 2022). Portanto, o principal nó central de seleção nesta rede é um complexo bioquímico unificado. As ASVs que se conectam a este módulo central representam a guilda especialista da floresta secundária. Elas não estão respondendo apenas ao pH, ou apenas ao carbono; elas estão respondendo a um ambiente

complexo e estável. As ASVs deste grupo compartilham afinidade com condições de estrutura e retenção hídrica do solo, onde ocorrem muitas ligações de ASVs de água disponível com porosidade total do solo, indicando uma adaptação à heterogeneidade física da matriz edáfica (Du et al., 2022)

Este padrão é a assinatura de um ecossistema K-estrategista maduro, onde a disponibilidade de nutrientes (Química) é controlada pela decomposição e ciclagem biológica, em vez de inputs exógenos, como visto em solos florestais ricos em matéria orgânica (Chen, J. et al., 2022; Liu, X. et al., 2024; Byers et al., 2025). As ASVs associadas a este complexo representam a guilda especialista, adaptada a um ambiente estável e de ciclagem de nutrientes. Secundariamente, um módulo Físico independente, centrado em Density e Clay, foi observado, indicando que a textura do solo atua como um filtro de habitat persistente, selecionando uma guilda distinta de ASVs adaptadas às propriedades estruturais do solo.

A análise topológica da rede de correlações nos mostra que em solos de Floresta Secundária os atributos físicos, especialmente porosidade (92 conexões), são fundamentais, determinando a estrutura microbiana ao controlar a disponibilidade de oxigênio e água. Isso corrobora a importância da estrutura física do solo na diversidade bacteriana, conforme Liu et al. (2024), que identificaram a porosidade como o principal preditor da diversidade bacteriana em solos florestais não degradados (Li et al., 2022; Li et al., 2024). Entre os fatores químicos, o magnésio emergiu como elemento-chave (60 conexões), funcionando como filtro fino que modula grupos bacterianos específicos, onde ocorre o aumento da riqueza bacteriana em solos com maiores níveis de Mg, mesmo em solos ácidos. Já os atributos biológicos, com destaque para urease (94 conexões), reforçam a centralidade da ciclagem de nitrogênio na comunidade microbiana. Além de que, enzimas como betaglucosidase e atividade respiratória, conectadas a processos de degradação de carbono e metabolismo aeróbico, sublinham o solo nativo como um ecossistema funcionalmente equilibrado e bioquimicamente integrado (Shange et al., 2012; Louisson et al., 2023; Liao et al., 2024).

Figura 21 - Rede de correlação de Spearman para o tratamento de Pastagem



Fonte: O autor.

A rede de correlação da PA (Figura 21) apresenta uma arquitetura única em comparação com a FS (integrada) e o PD (modularizada). A sua principal característica é a fragmentação ecológica, onde a rede se divide em dois componentes principais (módulos) totalmente desconectados, além de um nó isolado. Esta ausência de arestas (linhas) entre os dois módulos significa que os processos ecológicos e as guildas de ASVs associadas a cada módulo são desacoplados e independentes. As ASVs que prosperam sob as condições do módulo da esquerda não são as mesmas que prosperam sob as condições do módulo da direita; não há sobreposição de nicho positiva ($\rho > 0,4$).

Onde o módulo à esquerda, representa o núcleo funcional primário da ciclagem de nutrientes do solo da pastagem. É um sistema bioquimicamente acoplado, muito semelhante ao que vimos na FS. A atividade biológica (respiração, enzimas do ciclo C/S, glomalina) está

intrinsecamente ligada à química do solo (pH e disponibilidade de cátions como K e Mg) e aos recursos de base (Carbono Orgânico Total e água). E o atributo físico areia está neste módulo, sugerindo que este módulo metabólico opera predominantemente nas frações mais arenosas do solo (Xia et al., 2020). As ASVs conectadas aqui formam a guilda principal, adaptada à rizosfera das gramíneas e à ciclagem de nutrientes em solos bem drenados e quimicamente ativos (Huang et al., 2019; Costa et al., 2024; Xiang et al., 2025).

Enquanto o Módulo à direita, representa um módulo ecológico distinto, aparentemente independente do núcleo central da rede. A forte correlação positiva entre MBC (Biomassa Microbiana) e argila sugere que a manutenção da biomassa neste ambiente é mediada pela proteção física oferecida pelas partículas minerais. A fração argilosa favorece a formação de complexos organominerais e micro agregados estáveis, criando nichos protegidos que retêm umidade e substratos orgânicos, favorecendo a persistência de microrganismos mesmo sob condições limitantes. Neste contexto, o perfil enzimático associado (Urease e Fosfatase Ácida) indica uma atividade metabólica especializada na ciclagem de nutrientes adsorvidos a essa matriz mineral. Por fim, a resistência do solo (compactação) aparece como um nó isolado e negativamente conectado, atuando como um estressor físico que limita a conectividade da rede, em vez de um filtro que seleciona grupos específicos.

A fragmentação da rede sugere uma forte heterogeneidade espacial no solo da pastagem. A comunidade bacteriana responde a isso através de uma especialização de nicho, o consórcio funcional dominante (esquerda) domina as frações arenosas, enquanto uma guilda (direita) domina os micro-habitats de textura fina. As ASVs são tão especializadas que nenhuma delas prospera em ambos os ambientes o suficiente para criar uma correlação positiva entre si (Banerjee et al., 2016; Chen, W. et al., 2022; Yang et al., 2022; Chen et al., 2023; Baker et al., 2024; Cui et al., 2025). Juntos, esses resultados sustentam a ideia de que solos de pastagem intensivamente manejados formam redes microbianas altamente integradas, organizadas por uma interação sinérgica entre estrutura física, fertilidade química e atividade enzimática, garantindo produtividade e ciclagem contínua de nutrientes (Yang, et al., 2019; Wang, L. et al., 2023; Costa et al., 2024).

Os tratamentos de Preparo Convencional (PC) e Sistema Silvipastoril (SS) não produziram redes de correlação que satisfizesse os limiares estatísticos (Spearman $\rho > 0,4$; $p < 0,05$). Para o PC, a falta de correlações fortes e estáveis é atribuída à intensa perturbação física, que promove dinâmicas estocásticas e transitórias, impedindo a formação de guildas coesas, mesmo entre as suas 112 ASVs indicadoras.

Para o SS, a ausência de rede sugere um mecanismo oposto: uma heterogeneidade espacial em microescala tão alta (mosaico de nichos sob árvores, pasto e dejetos) que as correlações ASV-atributo, embora provavelmente fortes em micro-sítios, são diluídas e não detectadas no nível da amostragem composta (macro-ambiental). Isso é fortemente suportado pelo achado de que este sistema abriga a maior diversidade alfa (Shannon), refletindo sua alta complexidade de nicho. A análise integrada de módulos e nós centrais oferece uma visão holística, permitindo mapear tanto a organização modular da comunidade quanto os pontos centrais de influência funcional (Banerjee et al., 2016; Singavarapu et al., 2023; Byers et al., 2025; Cui et al., 2025; Idbella et al., 2025; Xiang et al., 2025).

4.6.1 Comparação entre redes - Nós Centrais

O sistema de PD caracteriza-se por uma rede em que os nós centrais são funcionalmente segregados. O sistema é controlado por múltiplos nós centrais que atuam em paralelo, mas de forma desacoplada. O nó central biológico controla a si mesmo, e a química (manejo) controla uma guilda separada. A rede da Floresta Secundária demonstra um mecanismo de controle oposto. Os nós centrais não são segregados. O sistema é controlado por um único e integrado nó central bioquímico. O nó central biológico e o químico estão acoplados, refletindo um sistema maduro e de ciclagem fechada, que operam de forma sinérgica, conferindo robustez e resiliência à comunidade mesmo sem a centralização de um único atributo (Chen, W et al., 2022; Yang et al., 2022; Chen et al., 2023; Baker et al., 2024; Cui et al., 2025).

A PA apresenta uma arquitetura de nós centrais mais complexa, caracterizada pela fragmentação. Ela não tem um único super nó central (como a FS) nem múltiplos nós centrais desacoplados (como o PD). Em vez disso, ela possui dois conjuntos de nós centrais independentes, refletindo uma especialização de nicho extrema. O controle é fragmentado. Existem duas guildas principais, cada uma com seu próprio conjunto de nós centrais independentes. Em comparação, o plantio apresenta menor redundância, com funções microbianas estruturadas por filtros texturais, químicos e biológicos claramente definidos, sugerindo eficiência funcional na ciclagem de carbono e na produtividade agrícola, porém com potencial vulnerabilidade caso um desses nós centrais seja impactado (Morrien et al., 2017; Cornell et al., 2023; Byers et al., 2024; Idbella et al., 2025). A PA é um sistema fragmentado, com múltiplos módulos de nós centrais (bioquímico e biofísico) operando de forma independente, refletindo uma alta especialização de nicho em um ambiente espacialmente heterogêneo.

Em suma, O PD é um sistema desacoplado, com nó central biológico, químico e físico atuando como filtros paralelos. A FS é um sistema acoplado, com um super nó central bioquímico unificado onde o biológico e o químico, refletem um ciclo fechado, sendo funcionalmente equilibrada (Chen, W et al., 2022; Du et al., 2022; Guseva et al., 2022; Liu et al., 2023; Liu, X. et al., 2024).

4.6.2 Comparação entre redes - Módulos

O sistema de módulo do PD exhibe uma estrutura segregada. A rede é altamente modular, particionada em três módulos funcionais distintos que operam em paralelo, mas de forma independente. Esta segregação é a assinatura de um ecossistema desacoplado (*decoupled*), típico de um agroecossistema de alta entrada (*high-input*). A Guilda de Resposta a *Inputs* (Química) existe para explorar os inputs externos de fertilizantes e calagem. A Guilda de Ciclagem Interna (Biológica) opera em paralelo, processando os atributos biológicos já existentes. A falta de conexões fortes entre esses dois módulos sugere que os processos são independentes, a fertilização química não alimenta diretamente o módulo biológico interno, e vice-versa. Segundo Yu et al. (2024), esse tipo de organização, comum em sistemas agrícolas intensivos, associa-se a uma forte especialização funcional de subcomunidades microbianas, com impacto direto no desempenho de processos como a ciclagem de nitrogênio e produtividade de culturas. Além disso, Singavaporu et al. (2023) mostram que a diversidade funcional das sub-redes pode ser reduzida em contextos de manejo intensivo, mesmo que a conectividade local aumente.

A FS apresenta uma arquitetura diametralmente oposta. Em vez de segregação, vemos integração e fusão dentre os módulos. Esta fusão define um ecossistema acoplado (*coupled*) e maduro (K-estrategista). Ela demonstra que a química não é um input externo, mas sim um resultado direto da biologia. O pH e a disponibilidade de K e Mg são controlados endogenamente pela atividade biológica (enzimas, decomposição da glomalina). Esta é a arquitetura de um sistema estável, e com ciclagem fechada de nutrientes. Esse padrão de alta modularidade é amplamente descrito em estudos como os de Szoboszlay e Tebbe (2021) e Kang et al. (2022), que evidenciam a presença de heterogeneidade funcional e modularidade elevada em solos florestais e sob diferentes tipos de vegetação arbórea.

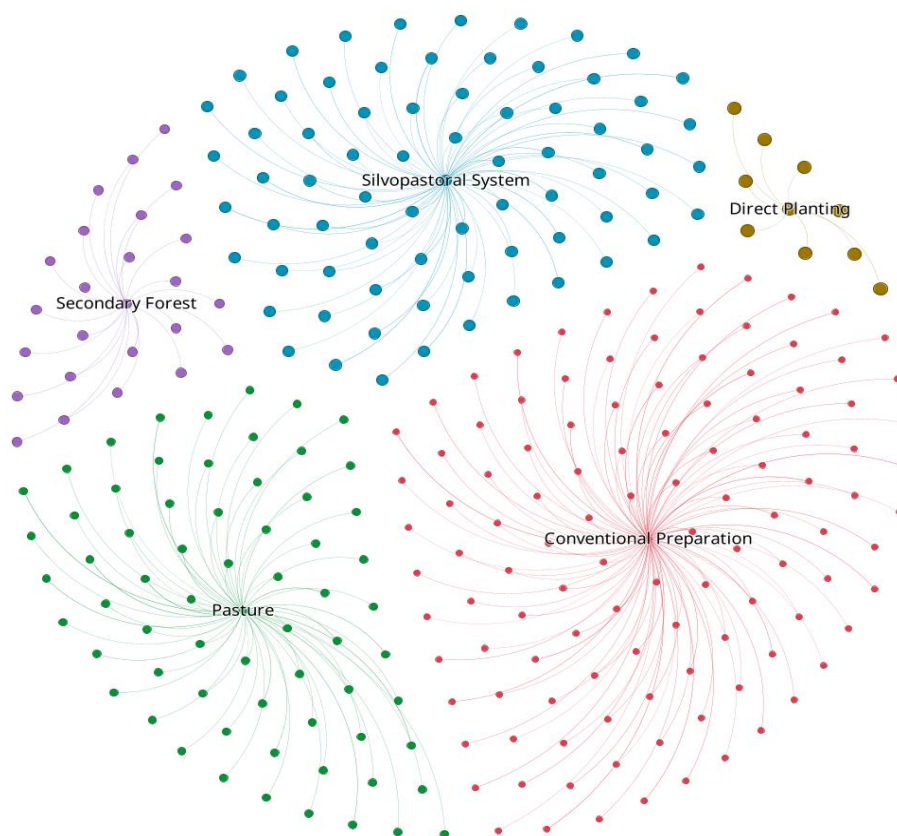
A PA apresenta a arquitetura mais complexa, definida não por módulos ou integração, mas por fragmentação. A rede inteira se quebra em componentes independentes, indicando uma especialização de nicho extrema. Este desacoplamento total entre os dois componentes é a

assinatura de um ambiente espacialmente heterogêneo. A Pastagem não é um único ambiente, mas pelo menos dois, sendo o primeiro o ambiente de ciclagem principal, e segundo um micro-habitat de textura fina (argila), que abriga uma guilda especialista completamente diferente, com seu próprio módulo biológico (MBC, Urease). O sistema se torna heterogêneo o que reflete na fragmentação de guildas ecologicamente independentes. Essa redução da multifuncionalidade e a reorganização das redes também foram observadas por Wang et al. (2023), que associaram tais mudanças a alterações nas condições ambientais em pastagens semiáridas, e por Cornell et al. (2023), que identificaram maior complexidade e estabilidade de rede em pastagens manejadas, mas menor modularidade. Sendo a transição de FS para agricultura, compatível com os padrões identificados por Idbella et al. (2025) em gradientes de uso do solo no sul da Itália, que demonstraram que mudanças no uso do solo reduzem a compartimentalização e aumentam a dependência de atributos centrais nas redes microbianas.

4.7 ANÁLISE REDE ESPÉCIE INDICADORA ATRAVÉS DO MÉTODO ÍNDICE DE VALOR INDICADOR COMPARADAS POR ÍNDICE LEVIN'S B

A Análise de Espécies Indicadoras (AEI) foi realizada para identificar os ASVs característicos de cada tipo de uso do solo. A análise traz diferenças quantitativas (número de ASVs indicadores) e qualitativas (identidade taxonômica), permitindo delinear como cada ambiente seleciona uma microbiota específica. Foram identificados 24 ASVs indicadores associados à FS, 9 ao PD, 66 à PA, 112 à PC e 71 à SS. No total, 282 ASVs (27,56% da biblioteca) foram considerados indicativos de um tratamento específico, enquanto 741 ASVs (72,43%) foram classificados como não indicadores. Em termos de abundância, os ASVs indicadores somaram 159.169 leituras, enquanto os não indicadores totalizaram 425.662 leituras. Na Figura 22, é apresentado, a AEI para os cinco tratamentos (Floresta Secundária, Pastagem, Plantio Direto, Preparo Convencional e Sistema Silvipastoril). Os nós centrais representam cada tratamento, e os nós menores conectados representam os ASVs que foram identificados como indicadores significativos ($p < 0,05$) para aquele ambiente específico.

Figura 22 - Análise de Espécies Indicadoras



Fonte: O autor.

O resultado desta análise é a quantidade de ASVs especialistas recrutados por cada tratamento (Figura 23). Não sendo apenas uma contagem, mas uma medida da intensidade e singularidade da pressão seletiva ambiental imposta por cada uso do solo. Onde no Preparo Convencional (PC) - 112 ASVs, o sistema recrutou o maior número de indicadores. Ele demonstra que o PC, caracterizado por intensa perturbação física (aração) e perturbação química (fertilização/pesticidas), não age como um filtro que empobrece a comunidade, mas sim como um filtro seletivo forte e único. Ele seleciona ativamente um grupo grande e altamente especializado de táxons (provavelmente r-estrategistas) que são unicamente adaptados para prosperar neste ambiente de alta perturbação e pulso de nutrientes.

Enquanto o Plantio Direto (PD) - 9 ASVs, recrutou o menor número de indicadores. Ao minimizar a perturbação física, atua como um ambiente de baixa pressão seletiva ou neutralidade ecológica. Ele não impõe um filtro ambiental forte o suficiente para selecionar um grupo único de especialistas. Em vez disso, o PD é o sistema que mais favorece o pool de generalistas, permitindo que a comunidade central (os 741 ASVs) domine. A presença de Acidobacteriota pode refletir períodos de menor atividade (como entressafas) ou práticas de

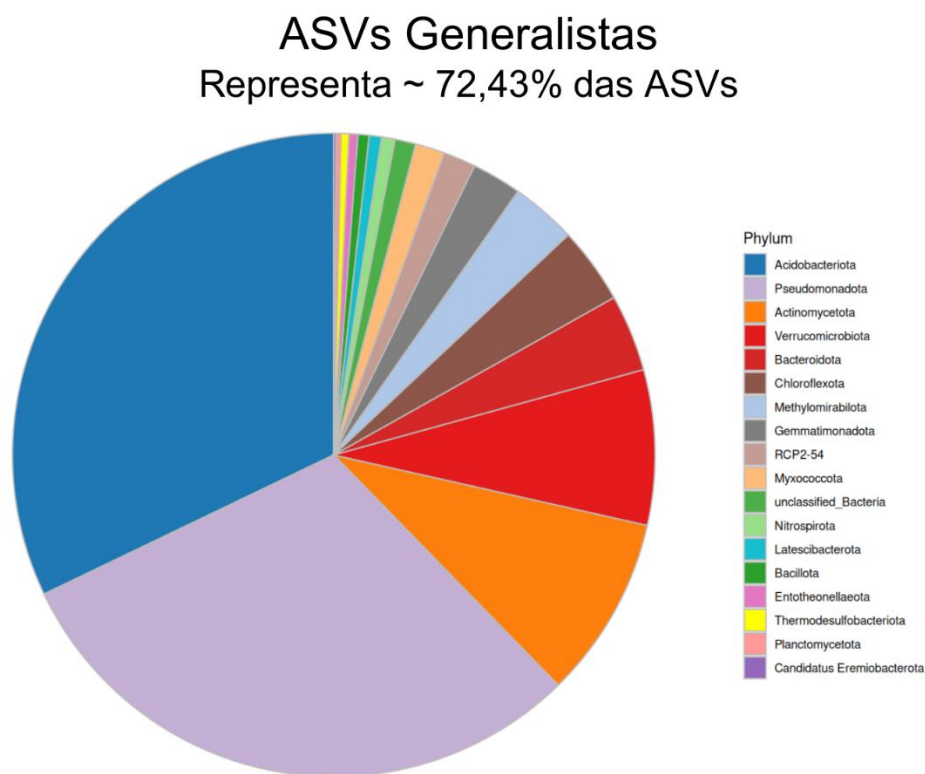
manejo específicas (por exemplo, ausência de animais ou aplicação reduzida de fertilizantes), o que o diferencia da pastagem (Shange et al., 2012; Wolinska et al., 2017; Fernandez-Gnecco et al., 2022; Chi et al., 2023; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024). A Floresta Secundária (FS) - 24 ASVs, selecionou um número moderado de especialistas. Consistente com um ecossistema maduro em clímax ou próximo a ele. O ambiente é estável, e a seleção é governada por filtros persistentes (ex: baixo pH, carbono recalcitrante), selecionando um grupo consistente, mas não numeroso, de especialistas (K-estrategistas), atuando como um sistema conservador de recursos (Shange et al., 2012; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024).

Já para Sistema Silvipastoril (SS - 71 ASVs) e a Pastagem (PA - 66 ASVs) recrutaram um número elevado e comparável de especialistas. Indicando que, ao contrário do PD, eles criam nichos ecológicos distintos. A introdução de animais (PA) ou a combinação de árvores e animais (SS) introduz novas fontes de carbono (dejetos, serrapilheira diversificada), gradientes de oxigênio (microsítios anaeróbicos) e interações rizosféricas (gramíneas vs. árvores) que selecionam ativamente um conjunto grande e específico de táxons. Na pastagem, essa composição indica um sistema de alta produtividade e rápida ciclagem de nutrientes, mas ecologicamente menos estável (Wolinska et al., 2017; Fernandez-Gnecco et al., 2022; Chi et al., 2023; Costa et al., 2024; Devkota et al., 2024).

ASVs Indicadoras



Figura 24 - Filos Espécies Não-Indicadoras em cada Tratamento



Fonte: O autor.

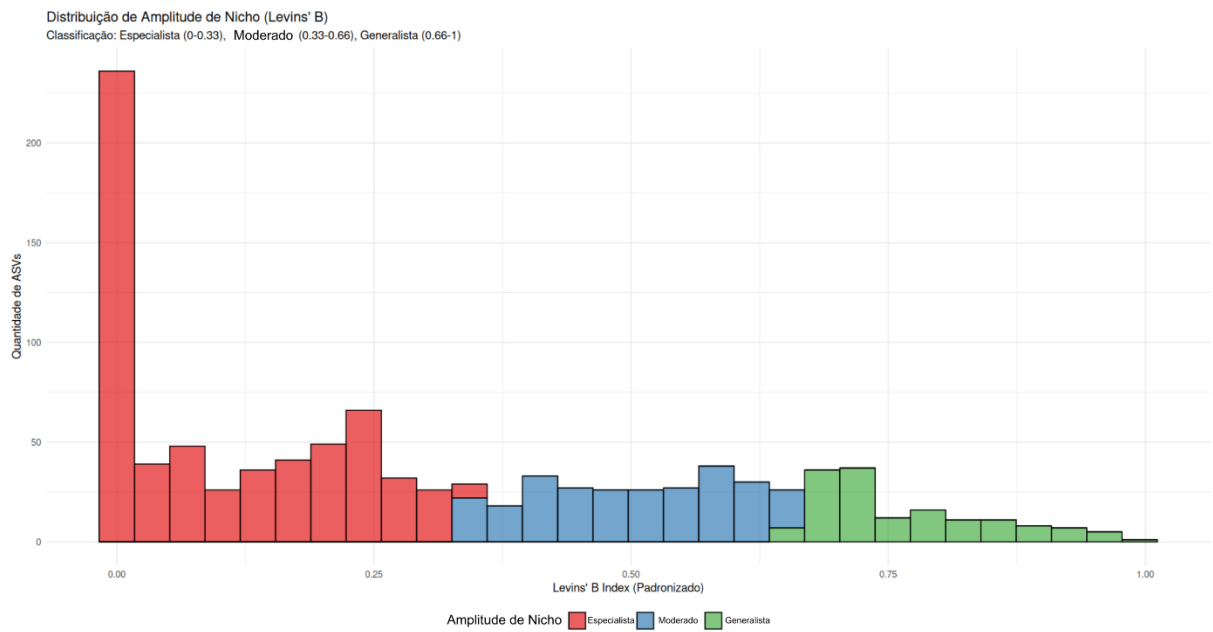
A análise de amplitude de nicho (Índice B de Levins), Figura 25 e Figura 26, quantificou a especialização de todas as 1023 ASVs. Os resultados demonstram que a comunidade é estruturalmente dominada por especialistas de nicho, com 606 ASVs classificadas, em oposição a apenas 151 generalistas. Foi identificada uma forte regra ecológica que governa a comunidade, a abundância total é positiva e significativamente correlacionada com a amplitude de nicho (Spearman's $\rho = 0,75$, $p < 2.2e-16$). Corroborado pelo fato de que as ASVs mais abundantes do estudo (ex: ASV_1 - 0,75, ASV_2 - 0,73;) são todas generalistas, representando o núcleo do microbioma (*core microbiome*) deste ecossistema. Esta análise também valida quantitativamente os resultados da AEI. As ASVs indicadoras são exceções a essa regra, onde as mesmas são especialistas. Por exemplo, a ASV_30 do filo Acidobacteriota (encontrada em PC) atingiu uma alta abundância (517) apesar de ter uma amplitude de nicho extremamente restrita (Levins' B = 0,22), confirmando seu papel como uma bioindicadora que prolifera em condições de manejo específicas, mas não em outras.

Figura 25 - Relação Amplitude Nicho Abundância Levins' B Index



Fonte: O autor.

Figura 26 - Distribuição Amplitude Nicho Levins' B Index



Fonte: O autor.

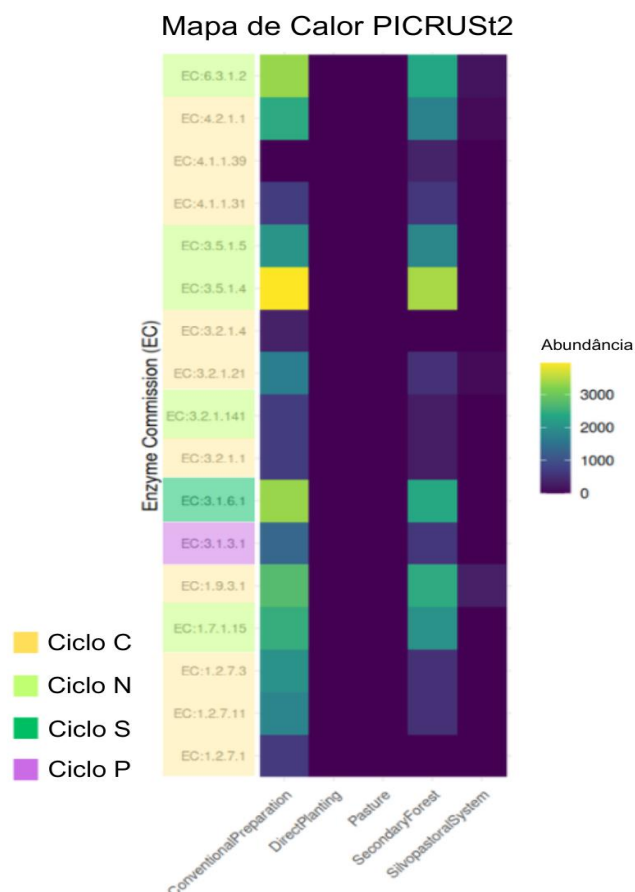
Explorando melhor os filos indicadores, os 24 especialistas da FS são quase que exclusivamente *Acidobacteriota* (aprox. 50%) e *Verrucomicrobiota* (aprox. 25%). Confirmando, em nível de ASV, que são os membros desses filos oligotróficos e K-estrategistas bio-indicadores de solos florestais estáveis e não perturbados. Apesar de seus números de indicadores diferentes, a identidade taxonômica dos especialistas do PD e do PC, é a mesma, onde ambos são dominados por *Pseudomonadota* e *Actinomycetota*. Indicando que, independentemente do nível de perturbação física, os ambientes agrícolas de plantio selecionam para os mesmos filos copiotróficos e r-estrategistas, adaptados a pulsos de nutrientes e degradação de matéria orgânica. Enquanto a Pastagem e o Sistema Silvipastoril selecionaram os pools de especialistas filogeneticamente mais diversos. Eles são os únicos tratamentos a recrutar indicadores significativos de filos como *Firmicutes* (provavelmente ligados a processos anaeróbicos/intestinais de dejetos animais), *Bacteroidota* e *Chloroflexi* (Fernandez-Gnecco et al., 2022; Chi et al., 2023; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024). Este resultado corrobora com a análise de diversidade alfa, onde a maior diversidade Shannon encontrada no Sistema Silvipastoril é diretamente explicada pelo fato de que a heterogeneidade estrutural e de substratos deste sistema (árvores + animais + gramíneas) cria um local de micro-nichos que suporta o conjunto de especialistas funcionalmente mais diversificado de todo o estudo.

4.8 ANÁLISE ABUNDÂNCIA RELATIVA DE VIAS METABÓLICAS INFERIDAS POR PICRUST2

A predição funcional da comunidade bacteriana foi realizada utilizando o algoritmo PICRUST2. Conforme ressaltado por Douglas et al. (2020), é importante notar que esta ferramenta não mensura a atividade enzimática *in situ*, mas sim a abundância de genes que codificam para funções específicas, permitindo inferir o potencial metabólico da comunidade. Os resultados para os diferentes usos do solo foram integrados em um único mapa de calor (Figura 27). Nesta representação gráfica, cada coluna corresponde a uma amostra e cada linha a uma função enzimática (EC - *Enzyme Commission numbers*). A legenda lateral categoriza as enzimas de acordo com os ciclos biogeoquímicos: Carbono (amarelo), Nitrogênio (verde claro), Enxofre (verde escuro) e Fósforo (roxo). A intensidade da cor na matriz (gradiente de amarelo claro a vermelho escuro) reflete a abundância relativa predita de cada gene funcional na comunidade.

A análise funcional do PICRUSt2 demonstra que a conversão do uso do solo induz um *trade-off* funcional que polariza a comunidade em dois extremos. A Floresta Secundária (FS), como sistema K-estrategista, exibe um potencial enzimático focado no Ciclo do Carbono, um investimento necessário para a decomposição de matéria orgânica recalcitrante, indicando que solos de floresta mantêm um perfil funcional estável e redundante, característica de ecossistemas em equilíbrio estrutural e funcional (Jensen et al., 2022; Singavarapu et al., 2023; Shu et al., 2023; Yang et al., 2024; Shi et al., 2024; Ma et al., 2024). A FS também apresenta um perfil funcional característico de um sistema equilibrado e metabolicamente eficiente, em que a via da respiração aeróbica predomina, sugerindo uma utilização eficaz do carbono como fonte de energia. A presença equilibrada de oxidoredutases associadas aos ciclos de nitrogênio e enxofre reforça a ideia de um solo maduro, com processos de ciclagem de nutrientes bem estabelecidos e diversificados, sem dependência excessiva de uma única via metabólica (Yu et al., 2024). Esse padrão funcional está alinhado com estudos comparativos que demonstram expressão de rotas metabólicas aeróbicas e de degradação de carbono em solos florestais conservados quando comparados com sistemas antropicamente manejados (Li et al., 2024; Yang et al., 2024).

Figura 27 - Mapa de calor da abundância relativa de vias metabólicas entre tratamentos



Fonte: O autor.

Em oposição, o Preparo Convencional (PC) representa o r-estrategista, exibindo o maior potencial funcional para os ciclos do Nitrogênio, Fósforo e Enxofre. Refletindo em um microbioma copiotrófico adaptado não à degradação complexa, mas ao consumo oportunista de pulsos de fertilizantes exógenos (Jensen et al., 2022; Chi et al., 2023; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024; Shi et al., 2024). Apesar disso, algumas enzimas apresentaram padrões consistentes de alta abundância relativa, como a EC:3.1.6.1 (N-acetilgalactosamina-6-sulfatase), EC:3.5.1.4 (Amidase) e EC:6.3.1.2 (Glutamina sintetase). A presença dessas enzimas indica uma comunidade microbiana funcionalmente voltada à ciclagem de nitrogênio e enxofre, refletindo a influência de matéria orgânica rica nesses elementos, oriunda da deposição de resíduos vegetais. A amidase e a glutamina sintetase são fundamentais, respectivamente, na liberação e assimilação de amônia, etapas centrais no metabolismo do nitrogênio. Já a sulfatase EC:3.1.6.1 participa da degradação de compostos sulfatados, frequentemente associados à matriz extracelular de resíduos orgânicos. O enriquecimento dessas mesmas rotas metabólicas em solos sob revolvimento do solo, indica

que essas adaptações são recorrentes em ambientes com alta carga de matéria orgânica nitrogenada (Das et al., 2022; Chi et al., 2023; Shu et al., 2023; Costa et al., 2024; Devkota et al., 2024; Le et al., 2025). A enzima EC:3.1.3.1, ligada ao ciclo do fósforo, também está presente, embora em menor abundância. Tal perfil reflete um microbioma agrícola adaptado para decompor exclusivamente os polímeros de carbono derivados da palhada e exsudatos radiculares, com menor diversidade funcional e maior eficiência em rotas específicas, provavelmente devido à seletividade imposta por práticas de manejo como fertilização e uso de defensivos (Jensen et al., 2022; Chi et al., 2023; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024; Shi et al., 2024).

Os demais sistemas de manejo (PA, PD, SS) posicionaram-se como funcionalmente intermediários, com valores de abundância extremamente baixos. Notavelmente, a conversão da FS para o PC força o microbioma a alterar sua função focada no Ciclo do Carbono, em favor de uma especialização em N/P/S, demonstrando um realinhamento funcional em resposta direta ao manejo do solo. Os resultados da análise funcional com PICRUSt2 demonstram que o uso do solo exerce influência determinística sobre a estrutura metabólica do microbioma do solo. Esses perfis são consistentes entre diferentes localidades e refletem adaptações funcionais aos respectivos regimes de manejo, com implicações claras para a função ecossistêmica do solo, como a ciclagem de nutrientes e a estabilidade bioquímica.

4.9 COMPARAÇÃO ENTRE ANÁLISES E TRATAMENTOS

Ao comparar as diferentes abordagens aplicadas, abundância relativa, diversidade alfa e beta, redes de correlação, espécies indicadoras, partição de variação (PV), algoritmos de AM e inferência funcional, demonstra-se um conjunto de padrões ecológicos consistentes que reforçam a influência determinística do uso do solo na estrutura e no funcionamento da comunidade bacteriana. Embora cada técnica capture dimensões distintas do sistema microbiano, todas convergem para o mesmo eixo interpretativo, de tal forma que se verifica que a conversão da Floresta Secundária (FS) para sistemas agrícolas promove mudanças na composição taxonômica, na especialização funcional e nos mecanismos de montagem da comunidade. De forma geral, ambientes conservados, como a FS, mantêm uma microbiota dominada por filos oligotróficos e altamente especializados, enquanto os sistemas agrícolas, especialmente Pastagem (PA) e Preparo Convencional (PC), favorecem grupos copiotróficos e comunidades funcionalmente reorganizadas, moldadas tanto pelos atributos biológicos do solo

quanto pelos insumos e perturbações físicas associadas ao manejo (Shange et al., 2012; Louisson et al., 2023; Domeignoz-Horta et al., 2024).

Na FS, essa convergência entre métodos fica evidente ao se observar que a comunidade bacteriana é marcada pela predominância de filos oligotróficos, como *Acidobacteriota* e *Verrucomicrobiota*, cuja elevada presença foi evidenciada tanto pela análise de abundância relativa quanto pelas espécies indicadoras e pela estrutura das redes de correlação. Esses filos, característicos de ambientes estáveis e de baixa perturbação, formam guildas especializadas associadas a atributos como acidez elevada, maior retenção de água e dinâmica de decomposição mais lenta (Shange et al., 2012; Louisson et al., 2023; Domeignoz-Horta et al., 2024). A própria rede da FS reflete esse padrão, exibindo um nó central bioquímico acoplado e altamente integrado, no qual as enzimas do ciclo do carbono, o pH e nutrientes como Mg e K atuam de maneira sinérgica na seleção das ASVs. A composição taxonômica, as métricas de diversidade e a organização topológica da rede convergem para caracterizar a FS como um sistema maduro, altamente modular e resiliente, confirmando que filtros ambientais persistentes, e não pulsos ou perturbações, são os principais responsáveis pela montagem dessa comunidade.

Na PA, o conjunto de análises indica para uma comunidade estruturalmente distinta da FS, caracterizada pela predominância de táxons copiotróficos e pela forte influência de atributos biológicos e químicos do solo. A abundância relativa revela um aumento expressivo de *Pseudomonadota*, acompanhado por *Bacteroidota* e *Firmicutes*, padrão compatível com ambientes enriquecidos por dejetos animais, maior disponibilidade de carbono lábil e zonas de compactação superficial (Wolinska et al., 2017; Fernandez-Gnecco et al., 2022; Chi et al., 2023). Esses grupos, frequentemente associados à ciclagem rápida de nutrientes, também emergem como espécies indicadoras relevantes e se organizam em redes fragmentadas, compostas por dois grandes módulos independentes, tal que temos um ligado à ciclagem ativa de nutrientes em frações arenosas e outro associado a micro-habitat de textura fina, dominado por ASVs especialistas correlacionados a urease, fosfatase ácida e maior teor de argila. Complementando esse cenário, a análise multivariada (PCoA/NMDS) posiciona as amostras de PA de forma consistente entre os sistemas agrícolas, separadas da FS e influenciadas por atributos como acidez potencial, carbono orgânico e glomalina. Assim, a Pastagem se destaca como um ecossistema produtivo, altamente dinâmico e menos estável, em que a especialização funcional e fragmentação espacial se combinam para criar uma comunidade com forte dependência das condições bioquímicas impostas pelo manejo.

No Plantio Direto (PD), os resultados evidenciam um sistema em que a comunidade bacteriana é estruturada por múltiplos filtros ambientais, porém mais equilibrados do que na PA e no PC. A abundância relativa indica uma composição intermediária, marcada pela presença simultânea de filos copiotróficos (como *Pseudomonadota*) e oligotróficos (como *Acidobacteriota*), refletindo o menor nível de perturbação física e a manutenção contínua de palhada na superfície do solo (Domeignoz-Horta et al., 2024; Liao et al., 2024). As espécies indicadoras são poucas, indicando baixa pressão seletiva e favorecimento do pool generalista (Kim et al., 2021; Zhao et al., 2022; Sun et al., 2024). Esse padrão também é observado na diversidade alfa, que apresenta variabilidade moderada sem extremos, e na diversidade beta, na qual o PD se agrupa claramente com outros usos agrícolas, mas mantém distância da FS. A topologia de rede reforça essa interpretação, onde os módulos são segregados e bem definidos, os quais são biológico, físico e químico respectivamente, sem a integração observada na FS nem a fragmentação extrema da pastagem. A separação entre esses módulos sugere que os processos ecológicos atuam em paralelo, com baixa redundância e interdependência entre si. As análises de PV e AM corroboram esse padrão, mostrando que o componente biológico é o preditor mais forte da abundância bacteriana, enquanto atributos químicos e físicos exercem influência secundária, porém consistente. Assim, o PD configura-se como um sistema de baixa seleção ambiental direta, sustentado por uma comunidade amplamente generalista, estável, porém com funcionalidade menos integrada do que a FS (Ding et al., 2013; Petersen et al., 2019; Bai et al., 2022; Yoon et al., 2024).

O Sistema Silvipastoril (SS) e o PC representam extremos contrastantes no gradiente agrícola, refletindo dois modos distintos de organização ecológica da comunidade bacteriana. No SS, observa-se uma microbiota altamente diversa (a maior diversidade alfa entre todos os usos do solo) sustentada por uma união de um micro-habitat derivado da coexistência entre árvores e gramíneas e da deposição de dejetos. Essa heterogeneidade espacial cria múltiplas oportunidades ecológicas, promovendo a coexistência de filos copiotróficos e oligotróficos, bem como um grande número, em relação aos demais, de ASVs indicadoras. Contudo, essa mesma complexidade espacial impede a formação de redes estáveis em escala macroambiental, resultando na ausência de correlações significativas nas análises de rede, indicando que os padrões ecológicos fortes provavelmente surgem apenas em micro-escalas dentro do sistema. O cenário observado no PC é antagônico aos sistemas conservados. Apesar de deter a maior quantidade de ASVs indicadoras, esse dado sugere um ambiente sob forte pressão seletiva, onde prevalecem taxons generalistas e oportunistas. A desestruturação das redes ecológicas e a baixa resiliência da comunidade são consequências diretas do manejo intensivo, que rompe a

integridade física dos agregados e favorece uma montagem de comunidade regida por distúrbios constantes. A abundância relativa revela o predomínio de *Pseudomonadota* e *Actinomycetota*, táxons adaptados a pulsos de fertilizantes e rápida mineralização, enquanto a PCoA posiciona o PC consistentemente afastado dos demais sistemas agrícolas, sugerindo que a perturbação mecânica atua como um filtro ecológico singular. Juntos, SS e PC ilustram dois caminhos opostos de montagem de comunidades, em que o primeiro promove diversidade por expansão de nichos, enquanto o segundo promove especialização por perturbações fortes e frequentes.

A integração entre as análises de ordenação (PCoA e NMDS) e os modelos explicativos (PV e AM) reforçam os mecanismos ecológicos identificados em cada uso do solo. As ordens multivariadas mostram que o uso do solo é o principal eixo de diferenciação da comunidade bacteriana, separando de forma consistente a FS dos sistemas agrícolas, enquanto fatores edafoclimáticos específicos, tais como o forte filtro ambiental de Castro, introduzem variações adicionais em cada grupo. A PV demonstra um funcionamento em conjunto dos atributos do solo, determinando a composição microbiana. Já os algoritmos de AM demonstram que o biológico é mais dominante, inclusive quando analisado táxon por táxon, indicando que a importância dos fatores biológicos não é apenas uma tendência generalizada na comunidade, mas uma regra estável que se repete na grande maioria dos filos e ASVs. Assim, tanto os métodos lineares quanto os não lineares indicam o mesmo resultado, que valida que a estrutura da comunidade microbiana é fortemente dirigida por filtros determinísticos e não apenas por processos estocásticos.

Ao integrar análises distintas demonstra que os padrões ecológicos identificados não são específicos de uma única abordagem. Tanto as ordenações não restritas (PCoA e NMDS) quanto as métricas de abundância relativa e diversidade revelam que a transição de FS para sistemas agrícolas desencadeia mudanças estruturais profundas na comunidade bacteriana, caracterizadas pelo aumento de filos copiotróficos, redução de oligotróficos e reorganização funcional associada ao manejo. Esses resultados refletem a lógica ecológica observada nas redes de correlação, nas quais sistemas preservados exibem nós centrais integrados e altas interdependências bioquímicas, enquanto sistemas agrícolas mostram fragmentação (PA), segregação funcional (PD) ou instabilidade (PC) (Kim et al., 2021; Zhao et al., 2022; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024; Sun et al., 2024). Em conjunto, esses padrões indicam que o uso do solo molda a comunidade microbiana por meio da alteração dos filtros ambientais dominantes, como da acidez e ciclagem de carbono na floresta, para a disponibilidade de nutrientes e suporte energético nos sistemas agrícolas, estabelecendo gradientes claros de especialização, resiliência e reorganização ecológica.

Os modelos preditivos e explicativos, PV e AM, fornecem um segundo eixo interpretativo que confirma a solidez dos padrões ecológicos detectados. A PV demonstrou que a maior parcela da variância explicada está associada a combinação dos atributos edáficos do solo. A inclusão de múltiplos algoritmos de AM reforça essa hierarquia: os três modelos baseados em árvores, *Random Forest (bagging)*, *XGBoost (boosting por gradiente extremo)* e *Gradient Boosting Machine (boosting clássico)*, convergem para o mesmo padrão, atribuindo maior importância aos atributos biológicos, seguido dos químicos e físicos para a grande maioria dos filos e ASVs. Além disso, os três métodos apresentaram altos valores de R^2 e de consistência interna, indicando que a estrutura da comunidade microbiana apresenta um sinal determinístico forte o suficiente para ser capturado mesmo por modelos de naturezas estatísticas diferentes. Essa convergência metodológica, entre técnicas lineares (PV) e não lineares (RF, XGBoost e GB), confirma que o padrão observado não é um artefato analítico, mas sim uma regularidade ecológica robusta que governa a montagem da comunidade bacteriana em todos os usos do solo avaliados.

A análise funcional inferida pelo PICRUSt2 complementa e fortalece os padrões ecológicos observados nas demais abordagens, demonstrando que as transições taxonômicas entre os usos do solo estão diretamente associadas a mudanças nas capacidades metabólicas da comunidade bacteriana. Estima-se que a FS apresenta enriquecimento em vias do Ciclo do Carbono, especialmente em oxidorreduções ligadas à respiração aeróbica e à degradação de compostos recalcitrantes, refletindo o perfil oligotrófico e de ciclagem lenta característico de sistemas maduros (Jensen et al., 2022; Singavarapu et al., 2023; Shu et al., 2023; Yang et al., 2024; Shi et al., 2024; Ma et al., 2024). Esse padrão funcional reforça o papel central de filos como *Acidobacteriota* e *Verrucomicrobiota*, que também emergem como indicadores e nós centrais ecológicos em ambientes menos perturbados. Em contraste, estima-se que sistemas agrícolas como PA e PC exibem forte aumento de vias associadas aos ciclos de Nitrogênio, Enxofre e Fósforo, particularmente amidases, ureases, sulfatases e enzimas de assimilação de amônia, que possuem perfis amplamente associados à disponibilidade de nutrientes lábeis e ao aporte frequente de insumos agrícolas (Das et al., 2022; Chi et al., 2023; Shu et al., 2023; Costa et al., 2024; Devkota et al., 2024; Le et al., 2025). Esse rearranjo funcional é coerente com a substituição de filos oligotróficos por copiotróficos, como *Pseudomonadota* e *Bacteroidota*, observada tanto nas análises de abundância relativa quanto nas espécies indicadoras (Kim et al., 2021; Zhao et al., 2022; Louisson et al., 2023; Domeignoz-Horta et al., 2024; Liao et al., 2024; Sun et al., 2024). Da mesma forma, a especialização funcional identificada pelo PICRUSt2 é compatível com a fragmentação ecológica detectada nas redes da PA e com a

instabilidade estrutural do PC, ambos dependentes de vias metabólicas rápidas e energeticamente eficientes. Assim, a integração da inferência funcional com os padrões taxonômicos, redes, ordenações e algoritmos de AM demonstra que o uso do solo não altera apenas a identidade dos microrganismos, mas reestrutura profundamente o repertório metabólico da microbiota, reforçando que a funcionalidade do ecossistema responde de maneira coerente às pressões seletivas impostas por cada sistema de manejo.

Em síntese, a integração de todas as abordagens empregadas, taxonômicas, funcionais, multivariadas, de redes ecológicas, espécies indicadoras, PV e algoritmos de AM, demonstra que o uso do solo atua como o principal modulador da estrutura e função da microbiota edáfica, enquanto fatores edafoclimáticos locais modulam finamente esses efeitos. A FS representa um sistema maduro e resiliente, sustentado por oligotróficos especializados, redes altamente integradas e um perfil funcional centrado na decomposição de carbono, coerente com sua estabilidade ecológica (Jensen et al., 2022; Louisson et al., 2023; Domeignoz-Horta et al., 2024). Em contraste, sistemas agrícolas como PA e PC exibem reorganizações profundas: maior abundância de copiotróficos, fragmentação ou instabilidade de redes, aumento de vias metabólicas rápidas e forte dependência de filtros biológicos associados à ciclagem de N, S e P (Chi et al., 2023; Costa et al., 2024; Devkota et al., 2024). Sistemas intermediários, como o PD e o SS, ocupam posições funcionais distintas, onde o PD apresenta uma comunidade generalista sustentada por filtros paralelos pouco integrados, enquanto o SS apresenta a maior diversidade e complexidade de nichos, ainda que sem formação de redes consolidadas em macroescala (Jesus et al., 2009; Ding et al., 2013; Liu et al., 2013; Petersen et al., 2019). A convergência entre PCoA, redes, AEI, PICRUST2, VarPart e algoritmos de AM reforça que estes padrões não são artefatos de métodos isolados, mas sinais ecológicos consistentes: a microbiota edáfica responde de forma determinística ao manejo e à estrutura física, química e biológica do solo. Assim, o capítulo evidencia que a conservação de ambientes florestais e a adoção de manejos complexos favorecem maior estabilidade e funcionalidade microbiana, enquanto manejos intensivos reduzem a redundância ecológica e reconfiguram profundamente as relações entre táxons, funções e atributos ambientais.

4.9.1 Conservação da Floresta Secundária próximo ao uso agrícola do solo

Os resultados deste estudo sugerem que a integração de sistemas mais complexos como o Sistema Silvipastoril e a conservação de Floresta Secundária (FS) adjacentes a áreas agrícolas, como pastagens e plantios, têm o potencial de desempenhar um papel fundamental na

manutenção da microbiota do solo e na promoção da resiliência funcional dos ecossistemas agrícolas. Autores como Goss-Souza et al. (2019) e Qu et al. (2024) exploram que a conversão de florestas para uso agrícola provoca reduções significativas na diversidade microbiana, homogeneização das comunidades e queda nas funções ecossistêmicas relacionadas ao carbono e ao ciclo de nutrientes. Nesse cenário, os dados apontam que os fragmentos de FS podem funcionar como reservatórios naturais de microrganismos especializados, incluindo táxons oligotróficos como *Acidobacteriota* e *Verrucomicrobiota*, melhorando a diversidade funcional e acelerando a recuperação de ecossistemas (Bell; Tylianakis, 2016; Delelegn et al., 2018; Goss-Souza et al., 2019; Chandel et al., 2022; Pratibha et al., 2023).

Além disso, a presença de remanescentes florestais contribui para aumentar a modularidade e complexidade das redes microbianas em áreas agricultáveis, favorecendo uma estrutura funcional mais robusta e multifuncional. Contudo, a partir do conceito ecologia de paisagens, o efeito de borda, a magnitude desse efeito ecossistêmico não é absoluta; ela é fortemente dependente de fatores espaciais, como o tamanho, a abundância e o grau de conectividade desses fragmentos florestais. Quando essas condições estruturais são favoráveis, análises como de Rijssel et al. (2022), Guo et al. (2023) e Liu J. et al. (2024) em ambientes de uso misto mostram que solos agrícolas próximos a florestas têm maior estabilidade e multifuncionalidade, derivadas de redes mais integradas e resilientes, especialmente em sistemas de agricultura de conservação e sistemas agroflorestais (Chandel et al., 2022; Pratibha et al., 2023; Álvarez et al., 2024; Sun et al., 2024; Santallanez-Arreola et al., 2025).

Sendo então, que conservar ou restaurar trechos de FS próximos a pastagens e plantios poderá não apenas preservar a biodiversidade taxonômica, mas também fortalece a composição biológica e funcional dos solos agrícolas. A diversificação vegetal promovida por sistemas consorciados e silvopastoris cria, através da heterogeneidade de exsudatos radiculares, microhabitats que favorecem a montagem de comunidades bacterianas distintas e funcionalmente complementares, superiores às encontradas em monoculturas, conforme destacado por Liu et al. (2024). O que fortalece serviços ecossistêmicos essenciais, tais como ciclagem de nutrientes, sequestro de carbono e resistência a perturbações, apoiando sistemas de produção agrícola mais sustentáveis e capazes de se auto regenerar após práticas intensivas (Liu, J. et al., 2023; Santallanez-Arreola et al., 2025).

5 CONCLUSÃO

O presente trabalho apresentou, de forma integrada, como diferentes abordagens taxonômicas, funcionais, multivariadas e computacionais podem revelar padrões consistentes sobre a estrutura e o funcionamento da comunidade bacteriana do solo sob distintos usos e manejos. A partir de análises de abundância relativa, diversidade alfa e beta, espécies indicadoras, redes de correlação ecológica, partição de variação, inferência funcional (PICRUSt2) e algoritmos de AM, foi possível demonstrar que o uso do solo é o principal determinante da composição microbiana, modulando tanto a identidade dos táxons quanto seu conjunto de funções metabólicas e suas interações ecológicas.

As análises taxonômicas evidenciaram que a conversão da Floresta Secundária para sistemas agrícolas promove profundas mudanças na assembleia bacteriana, com substituição de filos oligotróficos (*Acidobacteriota* e *Verrucomicrobiota*) por táxons copiotróficos associados à ciclagem rápida de nutrientes (*Pseudomonadota*, *Bacteroidota* e *Actinomycetota*) (Jensen et al., 2022; Shu et al., 2023; Yang et al., 2024; Ma et al., 2024). Esse padrão foi corroborado pela diversidade beta, na qual a PCoA evidenciou uma separação clara entre a Floresta Secundária e os usos agrícolas. Essa sensibilidade da estrutura da comunidade (diversidade beta), em contraste com a menor variação na riqueza total (alfa), reflete os padrões observados por Liu et al. (2024) em sistemas consorciados, onde a identidade funcional dos táxons mostrou-se mais responsiva ao manejo do que a biodiversidade bruta.

Aprofundando a análise dos vetores de mudança, observa-se que o manejo intensivo atua como um forte filtro ambiental (Kim et al., 2021; Zhao et al., 2022; Liao et al., 2024). Essa pressão seletiva é exacerbada pelo uso de insumos sintéticos; conforme destacado por Rijssel et al. (2022), a dependência química atua como uma perturbação crônica cuja interrupção é o gatilho primário para a reestruturação do microbioma em sistemas de transição, permitindo o retorno gradual de grupos suprimidos pelo estresse químico.

Na visão das interações ecológicas, as redes de correlação demonstraram que ambientes conservados mantêm maior integração bioquímica e estabilidade estrutural. Em contrapartida, sistemas agrícolas apresentaram fragmentação (Pastagem) ou instabilidade (Preparo Convencional), um comportamento típico de ambientes perturbados (Singavarapu et al., 2023; Shi et al., 2024). Essa perda de conectividade em sistemas simplificados corrobora os achados de Guo et al. (2023), que demonstraram que a diversidade vegetal atua como uma escada para a complexidade microbiana, onde sistemas com maior heterogeneidade vegetal sustentam redes de correlação mais robustas e interconectadas.

No âmbito funcional e de especialização, a AEI consolidou a Floresta Secundária como um ambiente dominado por especialistas, enquanto os sistemas manejados apresentaram indicadores relacionados à ciclagem acelerada e micro-habitats heterogêneos (Chi et al., 2023; Costa et al., 2024; Devkota et al., 2024). O PICRUS2 complementou essa visão, mostrando que a Floresta é marcada por vias de degradação de carbono recalcitrante, ao passo que os sistemas agrícolas exibiram enriquecimento em vias dos ciclos de Nitrogênio, Enxofre e Fósforo (amidases, ureases e sulfatases). Esse perfil metabólico reflete a adaptação a ambientes energeticamente dinâmicos (Cornell et al., 2023; Yu et al., 2024; Le et al., 2025), confirmando a plasticidade funcional da comunidade frente à disponibilidade de nutrientes.

Ao explorar com ferramentas de AM, identificou-se que as variáveis biológicas superaram as físico-químicas como preditores da comunidade (Louisson et al., 2023; Domeignoz-Horta et al., 2024). A alta capacidade preditiva desses modelos alinha-se a Hermans et al. (2020) e Wilhelm et al. (2022), comprovando que a assinatura biológica é consistente o suficiente para prever a qualidade do solo e o histórico de manejo com precisão. Ademais, essa consistência confirma a tese recente de Knight et al. (2024) de que microbiomas de solo respondem a perturbações de maneira determinística e previsível, validando o uso dessas ferramentas computacionais.

Do ponto de vista aplicado, o estudo destaca a relevância da conservação da Floresta Secundária e de sistemas como o Silvopastoril. Considerando o "efeito legado" descrito por Rijssel et al. (2022), onde o solo retém a memória do estresse passado por longos períodos, a preservação de refúgios biológicos torna-se mais eficiente do que a restauração tardia. A manutenção da complexidade vegetal favorece a estabilidade microbiana, reduz vulnerabilidades e aumenta a resiliência dos agroecossistemas (Álvarez et al., 2024; Sun et al., 2024; Santellanez-Arreola et al., 2025).

Apesar da robustez dos resultados, algumas limitações devem ser reconhecidas. O uso exclusivo do marcador 16S rRNA restringe a resolução taxonômica e funcional. Nesse sentido, a incorporação futura de métricas de biomassa molecular seria vital, uma vez que Maretto et al. (2023) revelaram que sistemas de manejo contrastantes podem diferir drasticamente na carga microbiana total (DNA total), uma dimensão de saúde do solo que complementa os dados de diversidade relativa aqui discutidos.

Em síntese, este trabalho demonstra que o componente biológico do solo não é apenas um reflexo passivo das condições ambientais, mas um módulo ativo da saúde edáfica. Nossos dados, em contraste com as evidências de Wilhelm et al. (2022), sugerem uma mudança de paradigma no monitoramento agrícola, tal que, a transição para diagnósticos integrativos

baseados no microbioma, onde oferece uma avaliação mais holística, antecipada e potencialmente mais custo-efetiva da sustentabilidade dos agroecossistemas.

REFERÊNCIAS

- ABELLAN-SCHNEYDER, I.; MATCHADO, M. S.; REITMEIER, S.; SOMMER, A.; SEWALD, Z.; BAUMBACH, J.; LIST, M.; NEUHAUS, K. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. **mSphere**, v. 6, n. 1, p. 1-22, 2021.
- ALONSO-SARRIA, F.; BLANCO-BERNARDEAU, A.; GOMARIZ-CASTILLO, F.; JIMÉNEZ-BASTIDA, H.; ROMERO-DIAZ, A. Estimation of soil properties using machine learning techniques to improve hydrological modeling in a semiarid environment: Campo de Cartagena (Spain). **Earth Science Informatics**, v. 18, n. 323, p. 1-24, 2025.
- ALVAREZ, V. E.; MUJTAR, V. A. E.; SALLES, J. F.; JIA, X.; CASTAN, E.; CARDOZO, A. G.; TITTONELL P. A. Micro-Environmental Variation in Soil Microbial Biodiversity in Forest Frontier Ecosystems—Implications for Sustainability Assessments. **Sustainability**, v. 16, n. 3, p. 1236, 2024.
- AMANN, R. I.; LUDWIG, W.; SCHIEFER, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. **Microbiological Reviews**, v. 59, n. 1, p. 143-169, 1995.
- ANDERSON M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). **Wiley StatsRef: Statistics Reference Online**. p. 1-15, 2017.
- AQEEL, M.; KHALID, N.; NOMAN, A.; RAN, J.; MANAN, A.; HOU, Q.; DONG, L.; SUN, Y.; DENG, Y.; LEE, S. S.; HU, W.; DENG, J. Interplay between edaphic and climatic factors unravels plant and microbial diversity along an altitudinal gradient. **Environmental Research**, v. 242, p. 117711, 2024.
- ARMALYTE, J.; SKERNISKYTE, J.; BAKIENE, E.; KRASAUSKAS, R.; SIUGZDINIENE, R.; KAREIVIENE, V.; KERZIENE, S.; KLIMIENE, I. Microbial diversity and antimicrobial resistance profile in microbiota from soils of conventional and organic farming systems. **Frontiers in Microbiology**, v. 10, p. 00892, 2019.
- ARUNRAT, N.; MHUANTONG, W.; SEREENONCHAI, S. Land-use legacies shape soil microbial communities and nutrient cycling functions in rotational shifting cultivation fields of Northern Thailand. **Microbial Ecology**, v. 88, art. 102, 2025.
- BAI, Z.; ZHENG, L.; BAI, Z.; JIA, A.; WANG, M. Long-term cultivation alter soil bacterial community in a forest-grassland transition zone. **Frontiers in Microbiology**, v. 13, p. 1001781, 2022.
- BAKER, N. R.; ZHALNINA, K.; YUAN, M.; FIRESTONE, M. K. Nutrient and moisture limitations reveal keystone metabolites linking rhizosphere metabolomes and microbiomes. **Proceedings of the National Academy of Sciences**, v. 121, n. 32, p. e2303439121, 2024.
- BANERJEE, S.; KIRKBY, C. A.; SCHMUTTER, D.; BISSETT, A.; KIRKEGAARD, J. A.; RICHARDSON, A. E. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. **Soil Biology and Biochemistry**, v. 97, p. 188-198, 2016.

BAY, S. K.; MCGEOCH, M. A.; WIELER, N.; PALMER, D. J.; BAKER, D. J.; CHOWN, S. L.; GREENING, C. Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution. **mSystems**, v. 5, n. 4, p. e00540-20, 2020.

BECKER, R. K.; GALVÃO, C. W.; ETTO, R. M.; HYEDA, D.; DE PAULA, A. L.; SCHIEBELBEIN, L. M.; GIAROLA, N. F. B. Assessing the impact of soil use and management systems on soil health in Southern Brazil. **Geoderma Regional**, v. 40, p. e00932, 2025.

BELL, T.; TYLIANAKIS, J. M. Microbes in the Anthropocene: spillover of agriculturally selected bacteria and their impact on natural ecosystems. **Proceedings of the Royal Society B**, 2016.

BENALCAZAR, P.; SEURADGE, B.; DIOCHON, A. C.; KOLKA, R. K.; PHILLIPS, L. A. Conversion of boreal forests to agricultural systems: soil microbial responses along a land-conversion chronosequence. **Environmental Microbiome**, v. 19, art. 32, 2024.

BEULE, L.; KARLOVSKY, P. Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. **PeerJ**, v. 8, p. e9593, 2020.

BOBULSKA, L.; ESPINDOLA, S. P.; COELHO, M. A.; FERREIRA, A. S. Impact of land use on soil function and bacterial community in the Brazilian savanna. **Anais da Academia Brasileira de Ciências**, v. 93, suppl. 3, p. e20201906, 2021.

BOLYEN, E.; RIDEOUT, J. R.; DILLON, M. R.; BOKULICH, N. A.; ABNET, C. C.; AL-GHALITH, G. A.; ALEXANDER, H.; ALM, E. J.; ARUMUGAM, M.; ASNICAR, F.; BAI, Y.; BISANZ, J. E.; BITTINGER, K.; BREJNROD, A.; BRISLAWN, C. J.; BROWN, C. T.; CALLAHAN, B. J.; CARABALLO-RODRIGUEZ, A. M.; CHASE, J.; COPE, E. K.; SILVA, R.; DIENER, C.; DORRESTEIN, P. C.; DOUGLAS, G. M.; DURALL, D. M.; DUVALLET, C.; EDWARDSON, C. F.; ERNST, M.; ESTAKI, M.; FOUQUIER, J.; GAUGLISTZ, J. M.; GIBBONS, S. M.; GIBSON, D. L.; GONZALEZ, A.; GORLICK, K.; GUO, J.; HILMANN, B.; HOLMES, S.; HOLSTE, H.; HUTTENHOWER, C.; HUTTLEY, G. A.; JANSSEN, S.; JARMUSCH, A. K.; JIANG, L.; KAEHLER, B. D.; KANG, K. B.; KEEFE, C. R.; KEIM, P.; KELLEY, S. T.; KNIGHTS, D.; KOESTER, I.; KOSCIOLEK, T.; KREPS, J.; LANGILLE, M. G. I.; LEE, J.; LEY, R.; LIU, Y.; LOFTFIELD, E.; LOZUPONE, C.; MAHER, M.; MAROTZ, C.; MARTIN, B. D.; MCDONALD, D.; MCIVER, L. J.; MELNIK, A. V.; METCALF, J. L.; MORGAN, S. C.; MORTON, J. T.; NAIMEY, A. T.; NAVAS-MOLINA, J. A.; NOTHIAS, L. F.; ORCHANIAN, S. B.; PEARSON, T.; PEOPLES, S. L.; PETRAS, D.; PREUSS, M. L.; PRUESSE, E.; RASMUSSEN, L. B.; RIVERS, A.; ROBESON, M. S.; ROSENTHAL, P.; TORRES, P. J.; TRINH, P.; TRIPATHI, A.; TURNBAUGH, P. J.; UL-HASAN, S.; HOOFT, J. J. J.; VARGAS, F.; VAZQUEZ-BAEZA, Y.; VOGTMANN, E.; HIPPEL, M.; WALTERS, W.; WAN, Y.; WANG, M.; WARREN, J.; WEBER, K. C.; WILLIAMSON, C. H. D.; WILLIS, A. D.; XU, Z. Z.; ZANEVELD, J. R.; ZHANG, Y.; ZHU, Q.; KNIGHT, R.; COPORASO, J. G. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. **Nature Biotechnology**, v. 37, p. 852-857, 2019.

BUKIN, Y. S.; GALACHYANTS, Y. P.; MOROZOV, I. V.; BUKIN, S. V.; ZAKHARENKO, A. S.; ZEMSKAYA, T. I. The effect of 16S rRNA region choice on bacterial community metabarcoding results. **Scientific Data**, v. 6, n. 1, p. 190007, 2019.

BÜNEMANN, E. K.; BONGIORNO, G.; BAI, Z.; CREAMER, R. E.; DEYN, G. D.; GOEDE, R.; FLESKENS, L.; GEISSEN, V.; KUYPER, T. W.; MÄDER, P.; PULLEMAN, M.; SUKKEL, W.; VAN GROENIGEN, J. W.; BRUSSAARD, L. Soil quality – A critical review. **Soil Biology and Biochemistry**, v. 120, p. 105-125, 2018.

BURNHAM, K. P.; OVERTON, W. S. Estimation of the size of a closed population when capture probabilities vary among animals. **Biometrika**, v. 65, n. 3, p. 625-633, 1978.

BURNHAM, K. P.; OVERTON, W. S. Robust estimation of population size when capture probabilities vary among animals. **Ecology**, v. 60, n. 5, p. 927-936, 1979.

BYERS, A. K.; WAKELIN, S. A.; CONDRON, L.; BLACK, A. Land use change disrupts the network complexity and stability of soil microbial carbon cycling genes across an agricultural mosaic landscape. **Microbial Ecology**, v. 87, art. 167, 2025.

CAIRES, S. A.; REINSCH, S.; SARAVANAKUMAR, D.; MARTIN, C. S.; WUDDIVIRA, M. N.; ZEBARTH, B. J.; KAYA, F.; LIU, M.; CHINTHALAPUDI, D. P. M.; SHANMUGAM, S. G.; HELGASON, B. Land use modifies the inherent effect of soil properties on soil bacterial communities in humid tropical watersheds. **Soil Systems**, v. 9, n. 4, p. 112, 2025.

CALLAHAN, B. J.; MCMURDIE, P. J.; ROSEN, M. J.; HAN, A. W.; JOHNSON, A. J. A.; HOLMES, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. **Nature Methods**, v. 13, p. 581-583, 2016.

CALLAHAN, B. J.; MCMURDIE, P. J.; HOLMES, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. **The ISME Journal**, v. 11, n. 12, p. 2639-2643, 2017.

CAO, Q.; SUN, X.; RAJESH, K.; CHALASANI, N.; GELOW, K.; KATZ, B.; SHAH, V. H.; SANYAL, A. J.; SMIRNOVA, E. Effects of rare microbiome taxa filtering on statistical analysis. **Frontiers in Microbiology**, v. 11, p. 607325, 2021.

CAPORASO, J. G.; LAUBER, C. L.; WALTERS, W. A.; KNIGHT, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. **Proceedings of the National Academy of Sciences**, v. 108, n. Suppl. 1, p. 4516-4522, 2011.

CARMO, K. B.; DIAS, R.; QUADROS, P. D.; BERBER, G. C. M.; BOURSCHEIDT, M. L. B.; NETO, A. L. F.; WEBER, O. L. S.; TRIPLETT, E. W.; FERREIRA, A. Assessment of soil bacterial communities in integrated crop production systems within the Amazon biome, Brazil: a comparative study. **Soil and Agricultural Microbiology**, v. 55, p. 2815-2825, 2024.

CHAO, A. Estimating the population size for capture-recapture data with unequal catchability. **Biometrics**, v. 43, n. 4, p. 783-791, 1987.

CHAO, A.; GOTELLI, N. J.; HSIEH, T. C.; SANDER, E. L.; MA, K. H.; COLWELL, R. K.; ELLISON, A. M. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. **Ecological Monographs**, v. 84, n. 1, p. 45-67, 2014.

CHANDEL, A.; MANN, R.; KAUR, J.; NORTON, S.; AUER, D.; EDWARDS, J.; SPANGENBERG, G.; SAWBRIDGE, T. The role of soil microbial diversity in the conservation of native seed bacterial microbiomes. **Microorganisms**, v. 10, n. 4, p. 750, 2022.

CHEN, J.; SONG, D.; LIU, D.; SUN, J.; WANG, X.; ZHOU, W.; LIANG, G. Soil aggregation shaped the distribution and interaction of bacterial-fungal community based on a 38-year fertilization experiment in China. **Frontiers in Microbiology**, v. 13, p. 824681, 2022.

CHEN, W.; WANG, J.; CHEN, X.; MENG, Z.; XU, R.; DUOJI, D.; ZHANG, J.; HE, J.; WANG, Z.; CHEN, J.; LIU, K.; HU, T.; ZHANG, Y. Soil microbial network complexity predicts ecosystem function along elevation gradients on the Tibetan Plateau. **Soil Biology and Biochemistry**, v. 172, p. 108766, 2022.

CHEN, J.; XIAO, Q.; XU, D.; LI, Z.; CHAO, L.; LI, X.; LIU, H.; WANG, P.; ZHENG, Y.; LIU, X.; QU, H.; BAO, Y. Soil microbial community composition and co-occurrence network responses to mild and severe disturbances in volcanic areas. **Science of the Total Environment**, v. 901, p. 165889, 2023.

CHI, Y.; SONG, S.; XIONG, K. Effects of different grassland use patterns on soil bacterial communities in the karst desertification areas. **Frontiers in Microbiology**, v. 14, p. 1208971, 2023.

COCK, P. J.; et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. **Bioinformatics**, v. 25, n. 11, p. 1422-1423, 2009.

COLWELL, R. K.; CHAO, A.; GOTELLI, N. J.; LIN, S.; MAO, C. X.; CHAZDON, R. L.; LONGINO, J. T. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. **Journal of Plant Ecology**, v. 5, n. 1, p. 3-21, 2012.

CORNELL, C. R.; ZHANG, Y.; NING, D.; XIAO, N.; WAGLE, P.; XIAO, X.; ZHOU, J. Land use conversion increases network complexity and stability of soil microbial communities in a temperate grassland. **The ISME Journal**, v. 17, p. 2210-2220, 2023.

COSTA, D. P.; SILVA, T. G. E.; ARAUJO, A. S. F.; PEREIRA, A. P. A.; MENDES, L. W.; BORGES, W. S.; FRANÇA, R. F.; SOUZA, C. A. F.; SILVA, B. A.; SILVA, R. O.; MEDEIROS, E. V. Soil fertility impact on recruitment and diversity of the soil microbiome in sub-humid tropical pastures in Northeastern Brazil. **Scientific Reports**, v. 14, p. 3919, 2024.

CUI, Y.; NING, Z.; LI, M.; QIN, X.; YUE, X.; CHEN, X.; ZHU, C.; SUN, H.; HUANG, Y. Microbial network-driven remediation of saline-alkali soils by salt-tolerant plants. **Frontiers in Microbiology**, v. 16, p. 1565399, 2025.

DAS, B. K.; ISHII, S.; ANTONY, L.; SMART, A. J.; SCARIA, J.; BROZEL, V. S. The microbial nitrogen cycling, bacterial community composition, and functional potential in a natural grassland are stable from breaking dormancy to being dormant again. **Microorganisms**, v. 10, n. 5, p. 923, 2022.

DAVIES, C. S.; WORSLEY, S. F.; MAHER, K. H.; KOMDEUR, J.; BURKE, T.; DUGDALE, H. L.; RICHARDSON, D. S. Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. **Microbiome**, v. 10, p. 41, 2022.

DELELEGN, Y. T.; PURAHONG, W.; SANDEN, H.; YITAFERU, B.; GODBOLD, D. L.; WUBET, T. Transition of Ethiopian highland forests to agriculture-dominated landscapes shifts the soil microbial community composition. **BMC Ecology**, v. 18, p. 58, 2018.

DEVKOTA, A. R.; WILSON, T.; KAUNDAL, A. Soil and root microbiome analysis and isolation of plant growth-promoting bacteria from hybrid buffaloberry (*Shepherdia utahensis* 'Torrey') across three locations. **Frontiers in Microbiology**, v. 15, p. 1396064, 2024.

DING, G.; PICENO, Y. M.; HEUER, H.; WEINERT, N.; DOHRMANN, A. B.; CARRILLO, A.; ANDERSEN, G. L.; CASTELLANOS, T.; TEBBE, C. C.; SMALLA, K. Changes of soil bacterial diversity as a consequence of agricultural land use in a semi-arid ecosystem. **PLoS One**, v. 8, n. 3, p. e59497, 2013.

DOMEIGNOZ-HORTA, L. A.; CAPPELLI, S. L.; SHRESTHA, R.; GERIN, S.; LOHILA, A. K.; HEINONSALO, J.; NELSON, D. B.; KAHMEN, A.; DUAN, P.; SEBAG, D.; VERRECCHIA, E.; LAINE, A. Plant diversity drives positive microbial associations in the rhizosphere enhancing carbon use efficiency in agricultural soils. **Nature Communications**, v. 15, p. 8065, 2024.

DORAN, J. W.; ZEISS, M. R. Soil health and sustainability: managing the biotic component of soil quality. **Applied Soil Ecology**, v. 15, n. 1, p. 3-11, 2000.

DOUGLAS, G. M.; MAFFEI, V. J.; ZANEVELD, J. R.; YURGEL, S. N.; BROWN, J. R.; TAYLOR, C. M.; HUTTENHOWER, C.; LANGILLE, M. G. I. PICRUSt2 for prediction of metagenome functions. **Nature Biotechnology**, v. 38, p. 685-688, 2020.

DU, M.; ZHANG, J.; WANG, G.; LIU, C.; WANG, Z. Response of bacterial community composition and co-occurrence network to straw and straw biochar incorporation. **Frontiers in Microbiology**, v. 13, p. 999399, 2022.

DUDEK, N. K.; CHAKHVADZE, M.; KOBAKHIDZE, S.; KANTIDZE, O.; GANKIN, Y. Supervised machine learning for microbiomics: bridging the gap between current and best practices. **Machine Learning with Applications**, v. 18, p. 100607, 2024.

EDGAR, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. **Nature Methods**, v. 10, n. 10, p. 996-998, 2013.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Diagnóstico das pastagens no Brasil**. Embrapa Amazônia Oriental, 2014. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/986147>>. Acesso em: 10 Outubro, 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Mapa de solos do estado do Paraná**. Embrapa Solos, n. 96, p. 600, 2007. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/339505>> Acesso em: 20 Abril, 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Plantio direto diminui emissão de carbono pelo solo**. Leona Landgraf e Rodrigo P. Barros, Embrapa Soja, 2009. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/18061875/plantio-direto-diminui-emissao-de-carbono-pelo-solo>>. Acesso em: 10 Outubro, 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Evolução tecnológica e arranjos produtivos de sistemas de integração lavoura-pecuária-floresta no Brasil**. Pesquisa Agropecuária Brasileira, v. 46, n. 10, p.1-12, 2011. Disponível em: <<http://www.alice.cnptia.embrapa.br/alice/handle/doc/912074>>. Acesso em: 10 Outubro, 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Regionalização dos testes de Valor de Cultivo e Uso e da indicação de cultivares de soja – Terceira Aproximação.** Embrapa Soja. n. 330, p. 69, 2015. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1022804>>. Acesso em: 20 Abril, 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Brasil lidera e é referência no desenvolvimento de tecnologias sustentáveis para produção de soja.** Lebna Landgraf, Embrapa Soja, 2023. Disponível em: <https://www.embrapa.br/tema-integracao-lavoura-pecuaria-floresta-ilpf/busca-de-noticias/-/noticia/81613580/brasil-e-referencia-no-desenvolvimento-de-tecnologias-sustentaveis-para-producao-de-soja?p_auth=e1hjtYH4>. Acesso em: 10 Outubro, 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Carbono da biomassa microbiana do solo em sistema de plantio direto com mais de 25 anos de implementação na Região Sul do Brasil.** v. 9, p. 130-133, 2025. Disponível em: <<http://www.alice.cnptia.embrapa.br/alice/handle/doc/1176667>>. Acesso em: 10 Outubro, 2025.

ETTO, R.; JESUS, E.; CRUZ, L.; SCHNEIDER, B.; TOMACHEWSKI, D.; URREA-VALENCIA, S.; et al. Influence of environmental factors on the tropical peatlands diazotrophic communities from the Southern Brazilian Atlantic Rain Forest. **Letters in Applied Microbiology**, v. 74, p. 543-554, 2022.

FAN, L.; SHAO, G.; PANG, Y.; DAI, H.; ZHANG, L.; YAN, P.; ZOU, Z.; ZHANG, Z.; XU, J.; ZAMANIAN, K.; DORODNIKOV, M.; LI, X.; GUI, H.; HAN, W. Enhanced soil quality after forest conversion to vegetable cropland and tea plantations has contrasting effects on soil microbial structure and functions. **Catena**, v. 211, p. 106029, 2022.

FERNANDEZ-GNECCO, G.; COVACEVICH, F.; CONSOLO, V. F.; BEHR, J. H.; SOMMERMAN, L.; MORADTALAB, N.; MACCARIO, L.; SØRENSEN, S. J.; DEUBEL, A.; SCHELLENBERG, I.; GEISTLINGER, J.; NEUMANN, G.; GROSCHE, R.; SMALLA, K.; BABIN, D. Effect of long-term agricultural management on the soil microbiota influenced by the time of soil sampling. **Frontiers in Soil Science**, v. 2, p. 837508, 2022.

FIERER, N. Embracing the unknown: disentangling the complexities of the soil microbiome. **Nature Reviews Microbiology**, v. 15, n. 10, p. 579-590, 2017.

FIERER, N.; WOOD, S. A.; MESQUITA, C. P. B. How microbes can, and cannot, be used to assess soil health. **Soil Biology and Biochemistry**, v. 153, p. 108111, 2021.

FUNNICELLI, M. I. G.; LIMA, N. S. M.; SARTINI, C. C. F.; LEMOS, E. G. M.; NETO, R. B. A.; SOUZA, H. A.; LIMA JUNIOR, J. O.; SAGRILO, E.; BLANCO, F. F.; ANDRADE, H. A. F.; SOUSA, D. C.; SILVA, M. L. N.; LEITE, L. F. C.; LIMA, P. S. C.; PINHEIRO, D. G. Revealing the bacteriome in crop–livestock–forest integration systems in the Cerrado of MATOPIBA, Brazil. **Forests**, v. 16, n. 4, p. 626, 2025.

GOSS-SOUZA, D.; MENDES, L. W.; BORGES, C. D.; RODRIGUES, J. L. M.; TSAI, S. M. Amazon forest-to-agriculture conversion alters rhizosphere microbiome composition while functions are kept. **FEMS Microbiology Ecology**, v. 95, n. 3, 2019.

GOSS-SOUZA, D.; TSAI, S. M.; RODRIGUES, J. L. M.; KLAUBERG-FILHO, O.; SOUSA, J. P.; BARETTA, D.; MENDES, L. W. Biogeographic responses and niche occupancy of microbial communities following long-term land-use change. **Antonie van Leeuwenhoek**, v. 115, p. 1129-1150, 2022.

GUO, Y.; YE, F.; SHENG, Q.; CLARK, T.; SAMUELS, D. C. Three-stage quality control strategies for DNA re-sequencing data. **Briefings in Bioinformatics**, v. 15, n. 6, p. 879-889, 2013.

GUO, J.; WEI, B.; LIU, J.; EISSENSTAT, D. M.; YU, S.; GONG, X.; WU, J.; HE, X.; YU, M. Linkages between plant community composition and soil microbial diversity in Masson pine forests. **Plants**, v. 12, n. 9, p. 1750, 2023.

GUSEVA, K.; DARCY, S.; SIMON, E.; ALTEIO, L. V.; MONTESINOS-NAVARRO, A.; KAISER, C. From diversity to complexity: microbial networks in soils. **Soil Biology and Biochemistry**, v. 169, p. 108604, 2022.

GWAK, H.; RHO, M. Data-driven modeling for species-level taxonomic assignment from 16S rRNA: application to human microbiomes. **Frontiers in Microbiology**, v. 11, 2020.

HUA, H.; SUI, X.; LIU, Y.; LIU, X.; CHANG, Q.; XU, R.; LI, M.; MU, L. Effects of land use type transformation on the structure and diversity of soil bacterial communities. **Life**, v. 14, n. 2, p. 252, 2024.

HERMANS, S. M.; BUCKLEY, H. L.; CASE, B. S.; CURRAN-COURNANE, F.; TAYLOR, M.; LEAR, G. Using soil bacterial communities to predict physico-chemical variables and soil quality. **Microbiome**, v. 8, p. 79, 2020.

HUANG, R.; MCGRATH, S. P.; HIRSCH, P. R.; CLARK, I. M.; STORKEY, J.; WU, L.; ZHOU, J.; LIANG, Y. Plant-microbe networks in soil are weakened by century-long use of inorganic fertilizers. **Microbial Biotechnology**, v. 12, n. 6, p. 1464-1475, 2019.

IDBELLA, M.; IACOMINO, G.; ABD-ELGAWAD, A. M.; BONANOMI, G. Soil microbial co-occurrence networks across climate and land use gradient in Southern Italy. **Environmental Microbiology Reports**, v. 17, n. 2, p. e70093, 2025.

JENSEN, G.; KROGSTAD, K.; RAZANEZHAD, F.; HUG, L. A. Microbial community compositional stability in agricultural soils during freeze-thaw and fertilizer stress. **Frontiers in Microbiology**, v. 10, 2022.

JESUS, E. C.; MARSH, T. L.; TIEDJE, J. M.; MOREIRA, F. M. S. Changes in land use alter the structure of bacterial communities in Western Amazon soils. **The ISME Journal**, v. 3, p. 1004-1011, 2009.

JIN, J.; LIU, X.; SHIROGUUCHI, K. Long journey of 16S rRNA-amplicon sequencing toward cell-based functional bacterial microbiota characterization. **iMetaOmics**, v. 1, n. 1, 2024.

KARIMI, B.; DEQUIEDT, S.; TERRAT, S.; JOLIVET, C.; ARROUAYS, D.; WINCKER, P.; CRUAUD, C.; BISPO, A.; PREVOST-BOURE, N. C.; RANJARD, L. Biogeography of soil bacterial networks along a gradient of cropping intensity. **Scientific Reports**, v. 9, p. 3812, 2019.

KANG, P.; PAN, Y.; YANG, P.; HU, J.; ZHAO, T.; ZHANG, Y.; DING, X.; YAN, X. A comparison of microbial composition under three tree ecosystems using the stochastic process and network complexity approaches. **Frontiers in Microbiology**, v. 13, p. 1018077, 2022.

KASCHUK, G.; ALBERTON, O.; HUNGRIA, M. Three decades of soil microbial biomass studies in Brazilian ecosystems: lessons learned about soil quality and indications for improving sustainability. **Soil Biology and Biochemistry**, v. 42, n. 1, p. 1-13, 2010.

KIBBLEWHITE, M. G.; RITZ, K.; SWIFT, M. J. Soil health in agricultural systems. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 363, n. 1492, p. 685-701, 2008.

KIM, H.; LEE, S.; JO, H. Y.; FINNERAN, K. T.; KWON, M. J. Diversity and composition of soil Acidobacteria and Proteobacteria communities as a bacterial indicator of past land-use change from forest to farmland. **Science of the Total Environment**, v. 797, p. 148944, 2021.

KNIGHT, R.; VRBANAC, A.; TAYLOR, B. C.; AKSENOV, A.; CALLEWAERT, C.; DEBELIUS, J.; GONZALEZ, A.; KOSCIOLEK, T.; MCCALL, L.; MCDONALD, D.; MELNIK, A. V.; MORTON, J. T.; NAVAS, J.; QUINN, R. A.; SANDERS, J. G.; SWAFFORD, A. D.; THOMPSON, L. R.; TRIPATHI, A.; XU, Z. Z.; ZANEVELD, J. R.; ZHU, Q.; CAPORASO, J. G.; DORRESTEIN, P. C. Best practices for analysing microbiomes. **Nature Reviews Microbiology**, v. 16, n. 7, p. 410-422, 2018.

KNIGHT, C. G.; NICOLITCH, O.; GRIFFITHS, R. I.; GOODALL, T.; JONES, B.; WESER, C.; et al. Soil microbiomes show consistent and predictable responses to extreme events. **Nature**, v. 636, n. 8043, p. 690-696, 2024.

KNIGHTS, D.; COSTELLO, E. K.; KNIGHT, R. Supervised classification of human microbiota. **FEMS Microbiology Reviews**, v. 35, n. 2, p. 343-359, 2011.

KWON, S.; PARK, S.; LEE, B.; YOON, S. In-depth analysis of interrelation between quality scores and real errors in Illumina reads. **IEEE Engineering in Medicine and Biology Society Conference**. p. 635-638, 2013.

LABOUYRIE, M.; BALLABIO, C.; ROMERO, F.; PANAGOS, P.; JONES, A.; SCHMID, M. W.; MIKRYOUKOV, V.; DULYA, O.; TEDERSOO, L.; BAHRAM, M.; LUGATO, E.; VAN DER HEIJDEN, M. G. A.; ORGIAZZI, A. Patterns in soil microbial diversity across Europe. **Nature Communications**, v. 14, p. 3311, 2023.

LAL, R. Restoring soil quality to mitigate soil degradation. **Sustainability**, v. 7, n. 5, p. 5875-5895, 2015.

LE, Q.; RU, M.; WEIBO, R.; MIN, Z.; WENTAO, M.; TING, Y.; YALING, L.; FENG, Y. Biological crusts increase soil microbial biomass and alter bacterial and fungal community compositions under pine restored over ten years in a Pisha sandstone soil. **Frontiers in Ecology and Evolution**, v. 13, 2025.

LEHMANN, J.; BOSSIO, D. A.; KÖGEL-KNABNER, I.; RILLIG, M. C. The concept and future of soil health. **Nature Reviews Earth & Environment**, v. 1, n. 10, p. 544-553, 2020.

LEI, C.; LU, T.; QIAN, H.; LIU, Y. Machine learning models reveal how biochar amendment affects soil microbial communities. **Biochar**, v. 5, p. 89, 2023.

LEI, J.; ZENG, C.; ZHANG, L.; WANG, X.; MA, C.; ZHOU, T.; LAFFITTE, B.; LUO, K.; YANG, Z.; TANG, X. Prediction of soil organic carbon stock combining Sentinel-1 and Sentinel-2 images in the Zoige Plateau, the northeastern Qinghai–Tibet Plateau. **Ecological Processes**, v. 13, p. 32, 2024.

LI, Z.; ZHANG, K.; QIU, L.; DING, S.; WANG, H.; LIU, Z.; ZHANG, M.; WEI, Z. Soil microbial co-occurrence patterns under controlled-release urea and fulvic acid applications. **Microorganisms**, v. 10, n. 9, p. 1823, 2022.

LI, Z.; KRAVCHENKO, A. N.; CUPPLES, A.; GUBER, A. K.; KUZYAKOV, Y.; ROBERTSON, G. P.; BLAGODATSKAYA, E. Composition and metabolism of microbial communities in soil pores. **Nature Communications**, v. 15, p. 3578, 2024.

LIAO, X.; CHEAN, Y.; HUANG, H.; ZHANG, H.; SU, Y.; ZHENG, D.; JIN, S. Response of soil microbial community in different forest management stages of Chinese fir plantation. **Forests**, v. 15, n. 7, p. 1107, 2024.

LIU, J.; ZHENG, C.; SONG, C.; GUO, S.; LIU, X.; WANG, G. Conversion from natural wetlands to paddy field alters the composition of soil bacterial communities in Sanjiang Plain, Northeast China. **Annals of Microbiology**, v. 64, p. 1395-1403, 2013.

LIU, S.; SUN, Y.; SHI, F.; LIU, Y.; WANG, F.; DONG, S.; LI, M. Composition and diversity of soil microbial community associated with land use types in the agro–pastoral area in the Upper Yellow River Basin. **Frontiers in Plant Science**, v. 13, p. 819661, 2022.

LIU, L.; MA, L.; ZHU, M.; LIU, B.; LIU, X.; SHI, Y. Rhizosphere microbial community assembly and association networks strongly differ based on vegetation type at a local environment scale. **Frontiers in Microbiology**, v. 14, p. 1129471, 2023.

LIU, J.; GUO, Y.; GU, H.; LIU, Z.; HU, X.; YU, Z.; LI, Y.; LI, L.; SUI, Y.; JIN, J.; LIU, X.; ADAMS, J. M.; WANG, G. Conversion of steppe to cropland increases spatial heterogeneity of soil functional genes. **The ISME Journal**, v. 17, p. 1872-1883, 2023.

LIU, W.; QIU, K.; XIE, Y.; HUANG, Y.; WANG, R.; LI, H.; MENG, W.; HE, Y.; LI, Y.; LI, H.; ZHAO, P.; YANG, Y. High-throughput absolute quantification sequencing reveals that a combination of leguminous shrubs is effective in driving soil bacterial diversity during the process of desertification reversal. **Microbial Ecology**, v. 86, p. 1145-1163, 2023.

LIU, J.; ZHANG, W.; TENH, C.; PANG, Z.; PENG, Y.; QIU, J.; LEI, J.; SU, X.; ZHU, W.; DING, C. Intercropping changed the soil microbial community composition but no significant effect on alpha diversity. **Frontiers in Microbiology**, v. 15, 2024.

LIU, X.; CHU, H.; GODOY, O.; DELGADO-BAQUERIZO, M. Positive associations fuel soil biodiversity and ecological networks worldwide. **Proceedings of the National Academy of Sciences**, v. 121, n. 6, p. e2308769121, 2024.

LOUISSEON, Z.; HERMANS, S. M.; BUCKLEY, H. L.; TAYLOR, B. S.; TAYLOR, M.; CURRAN-COURNANE, F.; LEAR, G. Land use modification causes slow, but predictable, change in soil microbial community composition and functional potential. **Environmental Microbiome**, v. 18, p. 30, 2023.

LORI, M.; KUNDEL, D.; MÄDER, P.; SINGH, A.; PATEL, D.; SISODIA, B. S.; RIAR, A.; KRAUSE, H. Organic farming systems improve soil quality and shape microbial communities across a cotton-based crop rotation in an Indian Vertisol. **FEMS Microbiology Ecology**, v. 100, n. 11, 2024.

LUAN, L.; JIANG, Y.; CHENG, M.; DINI-ANDREOTE, F.; SUI, Y.; XU, Q.; GEISEN, S.; SUN, B. Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. **Nature Communications**, v. 11, p. 6406, 2020.

LYNCH, M.; MARINOV, G. K. The bioenergetic costs of a gene. **Proceedings of the National Academy of Sciences**, v. 112, n. 51, p. 15690-15695, 2015.

MA, L.; ZHOU, G.; ZHANG, J.; JIA, Z.; ZOU, H.; CHEN, L.; ZHANG, C.; MA, D.; HAN, C.; DUAN, Y. Long-term conservation tillage enhances microbial carbon use efficiency by altering multitrophic interactions in soil. **Science of the Total Environment**, v. 915, p. 170018, 2024.

MA, Y.; SHEN, Y.; JIN, L.; TIAN, Y.; MA, H.; LAN, J.; FU, B. Vegetation–soil–microbiota dynamics across a 50-year reconstructed grassland chronosequence on the Loess Plateau of China. **PeerJ**, v. 12, p. e18723, 2024.

MARETTO, L.; DEB, S.; RAVI, S.; DE LUCIA, M. C.; BORELLA, M.; CAMPAGNA, G.; SQUARTINI, A.; CONCHERI, G.; NARDI, S.; STEVANATO, P. 16S metabarcoding, total soil DNA content, and functional bacterial genes quantification to characterize soils under long-term organic and conventional farming systems. **Chemical and Biological Technologies in Agriculture**, v. 10, p. 78, 2023.

MCMURDIE, P. J.; HOLMES, S. Waste not, want not: why rarefying microbiome data is inadmissible. **PLoS Computational Biology**, v. 10, p. e1003531, 2014.

MESZAROSOVA, L.; KUTÁKOVÁ, E.; KOHOUT, P.; MUNZBERGOVÁ, Z.; BALDRIAN, P. Closing the gap: examining the impact of source habitat proximity on plant and soil microbial communities in post-mining spoil heap succession. **Frontiers in Microbiology**, v. 15, p. 1416515, 2024.

MIZRAHI-MAN, O.; DAVENPORT, E. R.; GILAD, Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. **PLoS One**, v. 8, n. 1, p. e53608, 2013.

MO, L.; ZANELLA, A.; BOLZONELLA, C.; SQUARTINI, A.; XU, G.; BANAS, D.; ROSATTI, M.; LONGO, E.; PINDO, M.; CONCHERI, G.; FRITZ, I.; RANZANI, G.; BELLONZI, M.; CAMPAGNOLO, M.; CASAROTTO, D.; LONGO, M.; LINNYK, V.; IHLEIN, L.; YEOMANS, A. J. Effects of land conversion on soil microbial community structure and diversity in Songnen Plain, Northeast China. **Diversity**, v. 14, p. 638, 2022.

MORRIEN, E.; HANNULA, S. E.; SNOEK, L. B.; HELMSING, N. R.; ZWEERS, H.; HOLLANDER, M.; SOTO, R. L.; BOUFFAUD, M.; BUÉE, M.; DIMMERS, W.; DUYTS, H.; GEISEN, S.; GIRLANDA, M.; GRIFFITHS, R. I.; JØRGENSEN, H.; JENSEN, J.; PLASSART, P.; REDECKER, D.; SCHMELZ, R. M.; SCHMIDT, O.; THOMSON, B. C.; TISSERANT, E.; UROZ, S.; WINDING, A.; VAN DER PUTTEN, W. H. Soil networks become more connected and take up more carbon as nature restoration progresses. **Nature Communications**, v. 8, p. 14349, 2017.

NAVARRETE, A. A.; VENTURINI, A. M.; MEYER, K. M.; KLEIN, A. M.; TIEDJE, J. M.; BOHANNAN, B. J. M.; NÜSSLEIN, K.; TSAI, S. M.; RODRIGUES, J. L. M. Differential response of Acidobacteria subgroups to forest-to-pasture conversion and their biogeographic patterns in the western Brazilian Amazon. **Frontiers in Microbiology**, v. 6, p. 1443, 2015.

NARAYANA, N. K.; KINGERY, W. L.; SHANKLE, M. W.; SHANMUGAM, S. G. Differential response of soil microbial diversity and community composition influenced by cover crops and fertilizer treatments in a dryland soybean production system. **Agronomy**, v. 12, n. 3, p. 618, 2022.

NEARING, J. T.; DOUGLAS, G. M.; COMEAU, A. M.; LANGILLE, M. G. D. Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. **PeerJ**, v. 6, p. e5364, 2018.

NIKODEMOVA, M.; HOLZHAUSEN, E. A.; DEBLOIS, C. L.; BARNETT, J. H.; PEPPARD, P. E.; SUEN, G.; MALECKI, K. M. The effect of low-abundance OTU filtering methods on the reliability and variability of microbial composition assessed by 16S rRNA amplicon sequencing. **Frontiers in Cellular and Infection Microbiology**, v. 13, p. 103389, 2023.

PACE, N. R. A molecular view of microbial diversity and the biosphere. **Science**, v. 276, n. 5313, p. 734–740, 1997.

PETERSEN, I. A. B.; MEYER, K. M.; BOHANNAN, B. J. M. Meta-analysis reveals consistent bacterial responses to land use change across the tropics. **Frontiers in Ecology and Evolution**, v. 7, p. 391, 2019.

PICHLER, M.; COSKUN, Ö. K.; ORTEGA-ARBULÚ, A.-S.; CONCI, N.; WÖRHEIDE, G.; VARGAS, S.; ORSI, W. D. A 16S rRNA gene sequencing and analysis protocol for the Illumina MiniSeq platform. **MicrobiologyOpen**, v. 7, n. 6, p. e00611, 2018.

PORTILLO, M. C.; LEFF, J. W.; LAUBER, C. L.; FIERER, N. Cell size distributions of soil bacterial and archaeal taxa. **Applied and Environmental Microbiology**, v. 79, p. 7610–7617, 2013.

PRATIBHA, G.; MAJUNATH, M.; RAJU, B. M. K.; SRINIVAS, I.; RAO, K. V.; SHANKER, A. K.; PRASAD, J. V. N. S.; RAO, M. S.; KUNDU, S.; INDORIA, A. K.; KUMAR, U.; RAO, K. S.; ANNA, S.; RAO, C. S.; SINGH, V. K.; BISWAS, A. K.; CHAUDHARI, S. K. Soil bacterial community structure and functioning in a long-term conservation agriculture experiment under semi-arid rainfed production system. **Frontiers in Microbiology**, v. 14, p. 1102682, 2023.

PRODAN, A.; TREMAROLI, V.; BROLIN, H.; ZWINDERMAN, A. H.; NIEUWDORP, M.; LEVIN, E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. **PLoS One**, v. 15, n. 1, p. e0227434, 2020.

QI, W.; SONG, W.; QI, R.; LI, Y.; WANG, H.; LI, Y.; CHANG, Z. Land use types drive the distinct patterns of bacterial and fungal communities in soils from the semi-arid area. **Microbial Ecology**, v. 88, p. 43, 2025.

QU, X.; LI, X.; BARDGETT, R. D.; DELGADO-BAQUERIZO, M. Deforestation impacts soil biodiversity and ecosystem services worldwide. **Proceedings of the National Academy of Sciences**, v. 121, n. 13, p. e2318475121, 2024.

QUAST, C.; PRUESSE, E.; YILMAZ, P.; GERKEN, J.; SCHWEER, T.; YARZA, P.; PEPLIES, J.; GLÖCKNER, F. O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. **Nucleic Acids Research**, v. 41, n. D1, p. D590–D596, 2013.

RAMETTE, A. Multivariate analyses in microbial ecology. **FEMS Microbiology Ecology**, v. 62, n. 2, p. 142–160, 2007.

REIS, J. B. A.; OLIVEIRA, T. M. R.; SILVA, M. R. S. S.; LOPES, F. A. C.; PAULA, A. M.; PONTES, N. C.; VALE, H. M. M. Different land use systems in the Brazilian Cerrado and their effects on soil bacterial communities. **Microorganisms**, v. 13, n. 4, p. 804, 2025.

RIEHLE, K.; COARFA, C.; JACKSON, A.; MA, J.; TANDOM, A.; PAITHANKAR, S.; RAGHURAMAN, S.; MISTRETTA, T.; SAULNIER, D.; RAZA, S.; DIAZ, M. A.; SHUMLAN, R.; AAGAARD, K.; VERSALOVIC, J.; MILOSAVLJEVIC, A. The Genboree microbiome toolset and the analysis of 16S rRNA microbial sequences. **BMC Bioinformatics**, v. 13, p. S11, 2012.

RIJSSEL, S. Q.; VEEN, G. F.; KOORNEEF, G. J.; BAKX-SCHOTMAN, J. M. T.; HOOVEN, F. C.; GEISEN, S.; VAN DER PUTTEN, W. H. Soil microbial diversity and community composition during conversion from conventional to organic agriculture. **Molecular Ecology**, v. 31, n. 15, p. 4017–4030, 2022.

RISELY, A.; GILLINGHAM, M. A. F.; BÉCHET, A.; BRÄNDEL, S.; HENI, A. C.; HEURICH, M. Phylogeny- and abundance-based metrics allow for the consistent comparison of core gut microbiome diversity indices across host species. **Frontiers in Microbiology**, v. 12, p. 659918, 2021.

ROKINS, P. D.; GOPAL, N. O.; ANANDHAM, R.; SARASWATHI, R. The impact of different planting systems on the bacterial diversity of rice cultivated in saline soil based on 16S rRNA gene-based metagenomic insights. **Agriculture**, v. 12, n. 10, p. 1624, 2022.

ROOPASHREE, S.; ANITHA, J.; CHALLA, S.; MAHESH, T. R.; VENKATESAN, V. K.; GULUWADI, S. Mapping of soil suitability for medicinal plants using machine learning methods. **Scientific Reports**, v. 14, p. 3741, 2024.

SANTELLÁNEZ-ARREOLA, K.; MARTÍNEZ-GAMINO, M. A.; CONTANTE-GARCÍA, V.; ARREOLA-ÁVILA, J.; PEÑA, C. G.; SILLER-RODRÍGUEZ, Q. K.; TREJO-CALZADA, R.; NAVA-REYNA, E. Beyond soil health: the microbial implications of conservation agriculture. **Diversity**, v. 17, n. 2, p. 90, 2025.

SCHLÖTER, M.; NANNIPIERI, P.; SØRENSEN, S. J.; VAN ELSAS, J. D. Microbial indicators for soil quality. **Biology and Fertility of Soils**, v. 54, p. 1–10, 2018.

SCHLOSS, P. D.; WESTCOTT, S. L.; RYABIN, T.; HALL, J. R.; HARTMANN, M.; HOLLISTER, E. B.; LESNIEWSKI, R. A.; PARKS, D. H.; ROBINSON, C. J.; SAHL, J. W.; STRESS, B.; THALLINGER, G. G.; VAN HORN, D. J.; WEBER, C. F. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7537–7541, 2009.

SHANGE, R. S.; ANKUMAH, R. O.; IBEKWE, A. M.; ZABAWA, R.; DOWD, S. E. Distinct soil bacterial communities revealed under a diversely managed agroecosystem. **PLoS One**, v. 7, n. 7, p. e40338, 2012.

SHI, Y.; GAHAGAN, A. L.; MORRISON, M. J.; GREGORICH, E.; LAPEN, D. R.; CHEN, W. Stratified effects of tillage and crop rotations on soil microbes in carbon and nitrogen cycles at different soil depths in long-term corn, soybean, and wheat cultivation. **Microorganisms**, v. 12, n. 8, p. 1635, 2024.

SHU, X.; HU, Y.; LIU, W.; XIA, L.; ZHANG, Y.; ZHOU, W.; LIU, W.; ZHANG, Y. Linking between soil properties, bacterial communities, enzyme activities, and soil organic carbon mineralization under ecological restoration in an alpine degraded grassland. **Frontiers in Microbiology**, v. 14, 2023.

SIDDIQUI, N. Y.; MA, L.; BRUBAKER, L.; MAO, J.; HOFFMAN, C.; DAHL, E. M.; WANG, Z.; KARSTENS, L. Updating urinary microbiome analyses to enhance biologic interpretation. **Frontiers in Cellular and Infection Microbiology**, v. 12, p. 789439, 2022.

SILVA, D. P.; EPSTEIN, H. E.; VEGA THURBER, R. L. Best practices for generating and analyzing 16S rRNA amplicon data to track coral microbiome dynamics. **Frontiers in Microbiology**, v. 13, 2023.

SILVA, L. J.; FREITAS, I. C.; ABREU, L. H. G.; OLIVEIRA, D. M. S.; BARBOSA, D. L. A.; FIGUEREDO, C. C.; CAMILOTTI, V. L.; TEIXEIRA, M. V.; FRAZÃO, L. A. Converting low-productivity pasture to well-managed pasture and silvopastoral system cause relevant changes in soil chemical and microbiological characteristics. **Forests**, v. 15, n. 6, p. 1029, 2024.

SINCLAIR, L.; OSMAN, O. A.; BERTILSSON, S.; EILER, A. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. **PLoS One**, v. 10, n. 2, p. e0116955, 2015.

SINGAVARAPU, B.; DU, J.; BEUGNON, R.; CESARZ, S.; EISENHAEUER, N.; XUE, K.; WANG, Y.; BRUELHEIDE, H.; WUBET, T. Functional potential of soil microbial communities and their subcommunities varies with tree mycorrhizal type and tree diversity. **Microbiology Spectrum**, v. 11, n. 2, p. e04578-22, 2023.

SOARES, D. A.; LUPATINI, G. C.; SEKIYA, B. M. S.; MATEUS, G. P.; ANDRIGHETTO, C.; MODESTO, V. C.; SILVA, J. R.; LUZ, J. H. S.; GALINDO, F. S.; CRUSCIOL, C. A. C.; PAVINATO, P. S.; ANDREOTTI, M. Integrated crop–livestock systems as a strategy for the sustainable production of corn and soybean grain in tropical sandy soils. **Agronomy**, v. 14, n. 9, p. 2071, 2024.

SOUZA, I. M.; SAGRILO, E.; OLIVEIRA JUNIOR, J. O. L.; ARAÚJO, M. D. M.; MUNIZ, L. C.; COSTA, J. B.; POMPEU, R. C. F. F.; SOUSA, D. C.; ANDRADE, H. A. F.; OLIVEIRA NETO, E. D.; LEITE, L. F. C.; BLANCO, F. F.; LIMA, P. S. C.; SOUZA, H. A. Soil chemical quality in integrated production systems with the presence of native and exotic tree components in the Brazilian eastern Amazon. **Forests**, v. 15, n. 7, p. 1078, 2024.

SONG, T.; WANG, J.; XU, X.; SUN, C.; SUN, C.; CHEN, Z.; ZHANG, Y.; HAO, L. Microbial community and network differently reshaped by crushed straw or biochar incorporation and associated with nitrogen fertilizer level. **GCB Bioenergy**, v. 15, n. 10, p. 1255–1272, 2023.

STALEY, J. T.; KONOPKA, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. **Annual Review of Microbiology**, v. 39, n. 1, p. 321–346, 1985.

SUN, Z.; SUN, C.; ZHANG, T.; LIU, J.; WANG, X.; FENG, J.; LI, S.; TANG, S.; JIN, K. Soil microbial community variation among different land use types in the agro-pastoral ecotone of northern China is likely to be caused by anthropogenic activities. **Frontiers in Microbiology**, v. 15, p. 1390286, 2024.

SZOBOSZLAY, M.; DOHRMANN, A. B.; POEPLAU, C.; DON, A.; TEBBE, C. C. Impact of land-use change and soil organic carbon quality on microbial diversity in soils across Europe. **FEMS Microbiology Ecology**, v. 93, n. 12, 2017.

SZOBOSZLAY, M.; TEBBE, C. C. Hidden heterogeneity and co-occurrence networks of soil prokaryotic communities revealed at the scale of individual soil aggregates. **MicrobiologyOpen**, v. 10, n. 1, 2021.

TAN, G.; OPITZ, L.; SCHLAPBACH, R.; REHRAUER, H. Long fragments achieve lower base quality in Illumina paired-end sequencing. **Scientific Reports**, v. 9, p. 2856, 2019.

TAPAÇA, I. P. E.; OBIEZE, C. C.; PEREIRA, G. V. M.; FANGUEIRO, D.; COUTINHO, J.; FRAGA, I.; PARTELLI, F. L.; RAMALHO, J. C.; RIBEIRO-BARROS, A. I. Irradiance level and elevation shape the soil microbiome communities of Coffea arabica L. **Environmental Microbiome**, v. 19, p. 75, 2024.

TOMACHEWSKI, D.; GALVÃO, C. W.; DE CAMPOS JÚNIOR, A.; GUIMARÃES, A. M.; FERREIRA DA ROCHA, J. C.; ETTO, R. M. Ribopeaks: a web tool for bacterial classification through m/z data from ribosomal proteins. **Bioinformatics**, v. 34, n. 17, p. 3058–3060, 2018.

TOM, W. A.; JUDY, J. V.; KONONOFF, P. J.; FERNANDO, S. C. Influence of empirically derived filtering parameters, amplicon sequence variant, and operational taxonomic unit pipelines on assessing rumen microbial diversity. **Journal of Dairy Science**, v. 107, n. 11, p. 9209–9234, 2024.

VAUPEL, A.; KUSTERS, M.; TOUPS, J.; HERWIG, N.; BOSEL, B.; BEULE, L. Trees shape the soil microbiome of a temperate agrosilvopastoral and syntropic agroforestry system. **Scientific Reports**, v. 15, p. 1550, 2025.

WANG, Q.; WANG, W.; ZHONG, Z.; WANG, H.; FU, Y. Variation in glomalin in soil profiles and its association with climatic conditions, shelterbelt characteristics, and soil properties in poplar shelterbelts of Northeast China. **Journal of Forestry Research**, v. 31, p. 279–290, 2019.

WANG, C. H.; WU, L.; WANG, Z.; ALABADY, M. S.; PARSON, D.; MALUMO, Z.; FANKHAUSER, S. C. Characterizing changes in soil microbiome abundance and diversity due to different cover crop techniques. **PLoS One**, v. 15, n. 5, p. e0232453, 2020.

WANG, W.; SUN, Z.; MISHRA, S.; XIA, S.; LIN, L.; YANG, X. Body size determines multitrophic soil microbiota community assembly associated with soil and plant attributes in a tropical seasonal rainforest. **Molecular Ecology**, v. 32, p. 6294–6303, 2023.

WANG, L.; HAMEL, C.; LU, P.; WANG, J.; SUN, D.; WANG, Y.; LEE, S.; GAN, Y. Using enzyme activities as an indicator of soil fertility in grassland – an academic dilemma. **Frontiers in Plant Science**, v. 14, p. 1175946, 2023.

WANG, L.; LI, X. Soil microbial community and their relationship with soil properties across various landscapes in the Mu Us Desert. **Forests**, v. 14, n. 11, p. 2152, 2023.

WANG, X.; ZHANG, Q.; ZHANG, Z.; LI, W.; LIU, W.; XIAO, N.; LIU, H.; WANG, L.; LI, Z.; MA, J.; LIU, Q.; REN, C.; YANG, G.; ZHONG, Z.; HAN, X. Decreased soil multifunctionality is associated with altered microbial network properties under precipitation reduction in a semiarid grassland. **iMeta**, v. 2, n. 2, 2023.

WEISS, S.; XU, Z. Z.; PEDDADA, S.; AMIR, A.; BITTINGER, K.; GONZALEZ, A.; LOZUPONE, C.; ZANEVELD, J. R.; VÁZQUEZ-BAEZA, Y.; BIRMINGHAM, A.; HYDE, E. R.; KNIGHT, R. Normalization and microbial differential abundance strategies depend upon data characteristics. **Microbiome**, v. 5, p. 27, 2017.

WERNER, J. J.; ZHOU, D.; CAPORASO, J. G.; KNIGHT, R.; ANGENENT, L. T. Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. **The ISME Journal**, v. 6, n. 7, p. 1273–1276, 2011.

WILHELM, R. C.; VAN ES, H. M.; BUCKLEY, D. H. Predicting measures of soil health using the microbiome and supervised machine learning. **Soil Biology and Biochemistry**, v. 164, p. 108472, 2022.

WILLIS, A.; BUNGE, J. Estimating diversity via frequency ratios. **Biometrics**, v. 71, n. 4, p. 1042–1049, 2015.

WILLIS, A. D. Rarefaction, alpha diversity, and statistics. **Frontiers in Microbiology**, v. 10, 2019.

WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. **Proceedings of the National Academy of Sciences**, v. 74, n. 11, p. 5088–5090, 1977.

WOESE, C. R. Bacterial evolution. **Microbiological Reviews**, v. 51, n. 2, p. 221–271, 1987.

WOLINSKA, A.; FRAC, M.; OSZUST, K.; SZAFRANEK-NAKONIECZNA, A.; ZIELENKIEWICZ, U.; STEPNIEWSKA, Z. Microbial biodiversity of meadows under different modes of land use: catabolic and genetic fingerprinting. **World Journal of Microbiology and Biotechnology**, v. 33, p. 154, 2017.

WU, L.; WEN, C.; QIN, Y.; YIN, H.; TU, Q.; VAN NOSTRAND, J. D.; YUAN, T.; YUAN, M.; DENG, Y.; ZHOU, J. Phasing amplicon sequencing on Illumina MiSeq for robust environmental microbial community analysis. **BMC Microbiology**, v. 15, p. 125, 2015.

WU, S.; DENG, J.; YIN, Y.; QIN, S.; ZHU, W.; ZHOU, Y.; WANG, B.; RUAN, H.; JIN, L. Bacterial community changes associated with land use type in the forest montane region of Northeast China. **Forests**, v. 11, n. 1, p. 40, 2020.

XIA, Q.; RUFTY, T.; SHI, W. Soil microbial diversity and composition: links to soil texture and associated properties. **Soil Biology and Biochemistry**, v. 149, p. 107953, 2020.

XIANG, M.; LIANG, Z.; ZHANG, Y.; WU, J.; MA, T.; DUO, L.; ZHANG, X.; FU, G. Grazing intensity modifies soil microbial diversity and their co-occurrence networks in an alpine steppe, central Tibet. **Microorganisms**, v. 13, n. 1, p. 138, 2025.

YAN, T.; WANG, Z.; WANG, Z.; QIN, Y.; WANG, Z.; LI, S. Temporal variation in bacterial community characteristics shaped by habitat in a reservoir buffer strip in China. **Ecology and Evolution**, v. 15, n. 3, p. e70957, 2025.

YANG, Y.; ASHWORTH, A. J.; DEBRUYN, J. M.; WILLETT, C.; DURSO, L. M.; COOK, K.; MOORE, P. A.; OWENS, P. R. Soil bacterial biodiversity is driven by long-term pasture management, poultry litter, and cattle manure inputs. **PeerJ**, v. 7, p. e7839, 2019.

YANG, P.; YUAN, M.; QU, C.; LI, J.; HUA, R.; ZHAO, S.; ZHANG, B.; ZHANG, L. Metagenomic insight into the soil microbial functions across land uses. **Journal of Soils and Sediments**, v. 24, p. 3684–3693, 2024.

YANG, Y.; SHI, Y.; FANG, J.; CHU, H.; ADAMS, J. M. Soil microbial network complexity varies with pH as a continuum, not a threshold, across the North China Plain. **Frontiers in Microbiology**, v. 13, p. 895687, 2022.

YANG, Y.; CHEN, Y.; LI, Z.; ZHANG, Y.; LU, L. Microbial community and soil enzyme activities driving microbial metabolic efficiency patterns in riparian soils of the Three Gorges Reservoir. **Frontiers in Microbiology**, v. 14, p. 1108025, 2023.

YANG, G.; MA, Y.; MA, X.; WANG, X.; LU, C.; XU, W.; LUO, J.; GUO, D. Changes in soil organic carbon components and microbial community following spent mushroom substrate application. **Frontiers in Microbiology**, v. 15, 2024.

YOON, J.; ADHIKARI, M.; JEONG, S. S.; LEE, S. P.; KIM, H. S.; LEE, G. S.; PARK, D. H.; KIM, H.; YANG, J. E. Microbial diversity of soils under different land use and chemical conditions. **Applied Biological Chemistry**, v. 67, p. 111, 2024.

YARZA, P.; YILMAZ, P.; PRUESSE, E.; GLÖCKNER, F. O.; LUDWIG, W.; SCHLEIFER, K. H.; WHITMAN, W. B.; EUZÉBY, J.; AMANN, R.; ROSSELLÓ-MÓRA, R. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. **Nature Reviews Microbiology**, v. 12, n. 9, p. 635–645, 2014.

YU, S.; CHEN, X.; WANG, T.; YAO, S.; PENG, X. Linking main ecological clusters of soil bacterial–fungal networks and nitrogen cycling genes to crop yields under diverse cropping systems in the North China Plain. **Chemical and Biological Technologies in Agriculture**, v. 11, p. 96, 2024.

ZHANG, T.; LIU, Y.; SUI, X.; FREY, B.; SONG, F. Effects of land conversion on soil microbial community structure and diversity in Songnen Plain, Northeast China. **Sustainability**, v. 14, n. 17, p. 10767, 2022.

ZHAO, Y.; WANG, M.; YANG, Y.; SHANG, P.; ZHANG, W. Tillage promotes the migration and coexistence of bacterial communities from an agro-pastoral ecotone of Tibet. **Microorganisms**, v. 10, n. 6, p. 1206, 2022.

ZHENG, Q.; HU, Y.; ZHANG, S.; NOLL, L.; BÖCKLE, T.; DIETRICH, M.; HERBOLD, C. W.; EICHORST, S. A.; WOEBKEN, D.; RICHTER, A.; WANEK, W. Soil multifunctionality

is affected by the soil environment and by microbial community composition and diversity. **Soil Biology and Biochemistry**, v. 136, p. 107521, 2019.

ZHOU, R.; NG, S. K.; SUNG, J. J. Y.; GOH, W. W. B.; WONG, S. H. Data pre-processing for analyzing microbiome data – a mini review. **Computational and Structural Biotechnology Journal**, v. 21, p. 4804–4815, 2023.

ZHU, G.; LUAN, L.; ZHOU, S.; DINI-ANDREOTE, F.; BAHRAM, M.; YANG, Y.; GEISEN, S.; ZHENG, J.; WANG, S.; JIANG, Y. Body size mediates the functional potential of soil organisms by diversity and community assembly across soil aggregates. **Microbiological Research**, v. 282, n. 127669, p. 1-13, 2024.

ZINGER, L.; TABERLET, P.; SCHIMANN, H.; BONIN, A.; BOYER, F.; DE BARBA, M.; GAUCHER, P.; GIELLY, L.; GIGUET-COVEX, C.; IRIBAR, A.; RÉJOU-MÉCHAIN, M. Body size determines soil community assembly in a tropical forest. **Molecular Ecology**, v. 28, n. 3, p. 528–543, 2019.